Teacher and school effects on student achievement

by

Mario Fernando Sanclemente Villegas

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Economics

May 2016

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I have no words to express my gratitude to my supervisors *Professor Wiji Arulampalam* and *Professor Robin Naylor*. I feel very lucky of having the opportunity to live this life experience close to them. With them I have received the best academic training and research orientation. They have always encourage and support me unconditionally, sharing with me their knowledge and time with extreme generosity.

Thanks to the Warwick Economics Department, particularly to *Dr Roland Rathelot*, *Dr Clement de Chaisemartin*, and many others who gave me helpful comments in the CAGE-AMES Seminars and were willing to meet me and giving me their advice. Thanks to my friends at Warwick, they were my family here and made more enjoyable this process. I would especially want to thank to *Mrs Margaret Nash* for being always worried about us and keeping our days sweet and happy.

Thanks to my family; my parents *Fernando* and *Maria Cristina*, my brother *Daniel*, and my sister *Ximena*, because despite of the distance and time, they have always made me feel as if I have never left home. Thanks to my girlfriend, and future wife *Maria Teresa*, for helping me to achieve this dream.

Thanks to the Chilean Government for the financial support, provided by means of the **CONICYT BECAS CHILE Folio 72100370**.

Thanks to the **Chilean Ministry of Education (Mineduc)**, and particularly to the Study Centre (Centro de Estudios, Mineduc) for allowing me to use their data. Specially, to Mr. Felipe Flores who was always willing to help me and answering my enquires on time.

Declarations

I declare that the contents of this thesis are based on my own research in accordance with the regulations of Warwick University. The work in this thesis is original, unless where indicated by references. This thesis has not been submitted for examination at any other university.

Mario Sanclemente

September 2015

Abstract

This thesis investigates the topics related to the estimation of Teacher Effectiveness (TEs) and School Effectiveness (SEs). With regard to TEs, the focus is on the predicted teacher impact on improving academic achievement when a student from the median (or the mean) of the achievement distribution is exposed to a more effective teacher. Regarding SE estimates, we explore the evolution of SEs over a specific period of time. In both cases, we base our estimations on the Chilean education system, from which we have exclusive access to very rich datasets.

Our main objectives are summarised as follows: (i) to consistently estimate TEs and SEs using Value Added Models (VAMs), studying the most common estimation approaches used in the literature, and the required assumptions on which they are founded; (ii) to provide the first TE estimates for the Chilean educational context; and (iii) to investigate the evolution of SEs, identifying what factors are associated.

The thesis is organised into five chapters. In Chapter 2, we present a detailed review of TE estimations based on typical Value Added Models, which are derived from a general achievement function (GAF). We then discuss some estimation methodologies and the validations of the estimations found in the literature. In Chapter 3, we present the data and describe how it is organised, placing special emphasis on the selected sample cohorts and the performance measures used through the thesis. In Chapter 4, we test for evidence of non-random assignment of pupils to classrooms (or teachers) in the Chilean context, in order to examine the random assignment assumption imposed in most of the VAMs.

For Chapter 5 and Chapter 6, we choose the VAM that enable us to estimate TEs and SEs simultaneously. We employ the Maximum Likelihood estimation (MLE) methodology and obtain predictions of teacher and school effects from the estimated empirical Bayes (EB) distributions. In both chapters, we discuss the assumptions required to consistently estimate our TE and SE measures using this method. We usually conduct the estimations under two VAM specifications, one with a preset value of the persistence parameter λ, and another with an unrestricted value of λ.

The results suggest that teachers are more able to generate a larger impact on Maths than on Language scores. If a pupil from the median of the standardised examination scores distribution were exposed to 1 standard deviation (SD) more effective teacher, she will move up around 9 percentile positions in Language and 12 percentile positions in Maths, in terms of the pupils' ranking by subject. Regarding the SE estimates in the long run, we find that neither **downward** nor **upward** trajectories of SEs are explained by differences in observed characteristics, apart from pupil academic performance. We find evidence that trajectories of school

effectiveness are associated with the proportion of *High* (or *Low*) quality teachers, based on our estimated TEs.

We conclude that teachers are important in improving pupil academic performance, and that the level of teacher quality within schools is related to the stability and trajectories of school effectiveness in the long run.

Abbreviations

- **GS**: Cohort Specialists
- **CST**: California Standards Tests
- **CPF**:Cumulative Production Function
- **EB**: Empirical Bayes
- **FGLS**: Feasible Generalised Least Square
- **FE**: Fixed Effects
- **GS**: Grade Specialists
- **GAF**: General Achievement Function
- **HLM**: Hierarchical Linear Models
- **HP**: High Performance Counterfactuals
- **KS**: Kolgomorov-Smirnov test
- **LAUSD**: Los Angeles Unified School District
- **LP**: Low Performance Counterfactuals
- **MLE**: Maximum Likelihood Estimation
- **Mineduc**: Ministry of Education, Chile
- **Non-RA**: Non-Random Assignment Index
- **NTS**: National Teacher Statute
- **NBPTS**: National Board for Professional Teaching Standards

- **NTS**: National Teacher Statute
- **OLS**: Ordinary Least Squares
- **PSS**: Priority Scholar Subsidy
- **SRT**: Perfectly Sorted Counterfactuals
- **RDM**: Random Counterfactuals
- **RE**: Random Effects
- **RSP**: Reduced School Panel
- **SEs**: School Effects
- **SoE**: Sorting Evidence
- **Simce**: Standardised National Examination, Chile
- **SScr**: Simce Scores
- **LMrk**: Language Marks
- **MMrk**: Maths Marks
- **SPD**: Student Panel Dataset
- **SS**: Subject Specialist
- **TCAP**: Tennessee Comprehensive Assessment Program
- **TEs**: Teacher Effects
- **TVAAS**: Tennessee Value-Added Assessment System
- **VAMs**: Value-Added Models
- **WG**: Within-Group Estimator

Motivation

It is commonly accepted in the field of economic study that education plays an important role in determining economic productivity and growth. Regarding the accumulation of human capital, a key factor in the individual learning process is schooling. Schools are a very complex system where students, parents, teachers and principals mutually interact. Every component within the system contributes to create what we call the *educational context*, which is a fundamental determinant in the development of pupils performance, teachers skills and the whole education experience.

Different *educational contexts* may favour or hinder the pupil's learning process. Exposure to an unfavourable context is highly costly for both individual students and for wider society. Thus, it is important to understand how to shape the *educational context* so as to enhance, as far as possible, the individuals learning and the acquisition of human capital.

Education is a direct mechanism to address inequality issues because it may generate social and economic intergenerational mobility (Dearden et al. (1997)). Society, in general, benefits when education reaches a wider range of the population, especially less-advantaged people. Differences in quality of education may exacerbate current income and opportunity gaps. Blanden and Machin (2010) suggest that initial academic differences between pupils whose parents have high levels of education and pupils whose parents have low levels of education are likely to increase due to the impact of the ongoing exposure to a more advantageous environment provided by the highly educated parents.

Educational studies have proposed a huge variety of policies intended to improve educational quality, particularly for more disadvantaged schools. These policies include: investment in infrastructure, additional staff training, teacher specialisation in subjects or study areas, incentive pay schemes, and reduction in the number of pupils per class. These differences also generate inequalities, in terms of opportunities, for the most vulnerable children who are regularly associated with low quality schools.

Hence, we consider it extremely important to find out the most cost-effective solutions to enhance education quality in contexts where more is needed. The literature has not totally agreed which of these policies are more effective. Krueger and Whitmore (2001) suggest that the reduction from a medium class size (22-25 students) to a small class size (13-17 students) would boost achievement test scores by a standard deviation (SD) of 0.2 the equivalent of an increase in scores of 5 to 7 percentage points. However, findings regarding the effects of class size on achievement are not uniform. Hoxby (2000) finds no significant impact on text scores when reducing class size, while Rockoff (2004) reviews the literature of class size effects estimation obtained from field experiments and suggests a circumspect interpretation of causal effects conclusions. Hence, these estimations may be misleading, depending on the context. For example, in ChileUrquiola and Verhoogen (2009) finds that a smaller class size is not necessarily associated with higher academic performance.

Our results from Chapters 5 and 6 suggest a similar magnitude of gains in academic achievement when a pupil from the mean of the performance distribution is exposed to a teacher who is one standard deviation more effective. However, we must be cautious when interpreting these results with design of policy in mind; the magnitude of the effectiveness, and potential differences among specific programmes might be difficult to compare.

This paper is organised as follows: In Chapter 1, we offer an introductory description of the key characteristics of the Chilean educational system, the setting for our analysis. The Chilean school system provides a valuable context for studying educational issues; Chile was one of the first countries to implement a national school voucher system, and Chile also has a very detailed data set available. The Ministry of Education (Mineduc) manages students, teachers and school registers making possible to keep tracking them on yearly basis.

In Chapter 2, we present a detailed review of Teacher Effect (TE) estimations based on common Value Added Models (VAMs) found in the literature. Using a general cumulative education production function, we construct a general achievement function (GAF); from this function we are able to derive four VAMs. We discuss the required assumptions and typical estimation methodologies for these models.

Chapter 3 introduces the multiple sources of the Chilean school system data set on which we base our investigations throughout the thesis. We describe the available performance measures, and we discuss how we handle cases in which some availability constraints surface.

In Chapter 4, we test for students-to-teacher non-random assignment within schools in the Chilean school system. This is the most important assumption on which most VAMs are founded. We focus on the potential student grouping in 4^{th} and 5^{th} grades based on previous academic achievement. Here, we use reduced mini panels that allow us to track students over an appropriate period of time.

Chapter 5 presents our estimation methodology of teacher effects for the Chilean school system. We describe the strategy to simultaneously estimate not only teacher effects but also school effects, and unobserved student heterogeneities. To our knowledge these are the first teacher Value Added measures for the Chilean school system, and they provide an interesting comparison for the estimations undertaken in different institutional environments and technical contexts of the existing literature. Although we obtain teacher and school effects measures, in this chapter we mainly focus on the predictive differences of teacher's impact on pupil academic performance under different VAM estimations.

In Chapter 6, we expand our analysis by examining school effectiveness in the long run, and the composition of teacher type within schools. Moreover, we are particularly interested in the evolution of school effectiveness, and the factors that might explain positive or negative trajectories in the school quality classification over the years.

Chapter 7 presents our conclusions, highlights key findings, and offers further remarks.

Chapter 1

The Chilean school system

The Chilean school system is composed of a pre-school stage (nursery to kindergarten), leading to primary and secondary schools. Schools can provide education from pre-school levels (e.g. pre-kindergarten and kindergarten) through secondary schools. Most of the schools provide both primary and secondary education, although some offer either primary or secondary only.

Secondary education has been mandatory since 2003 and from 2014 the pre-school stage became compulsory; thus pupils start school around the age of 3. Therefore, the Chilean school system ensures universal and free access education from age 3 to approximately age 18.[1]

The pupils age and the corresponding grade are presented in Table 1.1, where we show how primary education consists of eight grades, and secondary education consists of four grades. These 12 years are known as the *KS-12* system.

Schools can be either state owned or private institutions, but irrespective of this distinction, all schools must teach minimum standards as set out in the national curriculum , which is designed by the Ministry of Education (Mineduc).

The following section describes the institutional organisation of the Chilean school system for the last 30 years.

1.1 School system organisation

Among developing countries, Chile was the first to introduce a school voucher programme. Chile is also a pioneer in implementing vouchers at a national level.

[1]These policies have implied an increase the government's budget on education. Some evidence of that is shown in Cox (2004): "Public expenditure on education in Chile went from 2.4% of Gross Domestic Product in 1990 to 4.4% in 2001, rising on average more than 6% yearly. Education's prime importance to individuals and government was clearest in the magnitude of the new resources invested in this sector, as Table 5 reveals. Public expenditure in education almost tripled, going from USD 907.8 million to USD 3.017 billion (in constant dollars) from 1990 to 2002; per student spending rose accordingly".

Table 1.1: Educational levels and pupils' age

	Grade	Age
Pre-School		
	Day-nursery	0-2
	Pre-kinder	3-4
	Kindergarten	5-6
Primary School		
	1	6-7
	2	7-8
	3	8-9
	4	9-10
	5	10-11
	6	11-12
	7	12-13
	8	13-14
Secondary School		
	9	14-15
	10	15-16
	11	16-17
	12	17-18

Source: Ministry of Education, Chile

By the early 1980s, only the Netherlands had a nationwide school voucher system; subsequently Chile and New Zealand adopted a national voucher system, and they were followed later by Sweden and Hungary in the early 1990s (Patrinos (2000); Ladd (2002)). In other countries, such as the United States, from 1990s, several school voucher programmes were designed and implemented at state or city levels. Hence, the Chilean education system, with its universal voucher programme, has been under the gaze of policymakers and academics around the world.

Generally speaking, the voucher consists of a fixed amount of money per pupil enrolled, paid directly to schools. Voucher supporters argued that the system would promote private schools in order to provide improved educational services to a wider spectrum of students. In theory, on the supply side, existing and new schools would provide a *good quality* education in order to compete for pupil enrolment (Friedman (1955, 1997)). On the demand side, parents would be able to choose between Municipal and Private schools in order to find the best feasible option in terms of quality, costs and accessibility.

1.1.1 The supply side

In Chile, the voucher programme was part of larger educational reform introduced in 1981. In the wake of the implementation of the rReform act, the system has since been composed of three types of schools: (i) *Municipal schools*, are state-owned schools, but with a decentralised local (municipal) administration. The most important decisions regarding allocation of municipal resources are made by mayors and their councils, while the income received from the vouchers is managed by the school's principal; (ii) *Private Voucher schools*, they are privately owned and managed establishments, which receive public funding via vouchers per pupil enrolled. They can also raise extra funding by charging parents a limited amount of fee, private donations, or via other targeted public educational programmes; (iii) *Unsubsidised Private* schools are privately owned and administered establishments. They do not apply for any public funds.

The three types of schools are different in significant ways when it comes to the rules of governance. *Municipal schools* face more constraints in terms of student selection and staff-hiring decisions. In theory, *Municipal schools* are not allowed to select students, and their staff contracts are ruled by the National Teacher Statute (NTS), leaving principals with less flexibility to control salaries, or to offer other types of pay schemes. In contrast, the other two types of schools; (*Private Voucher* and *Unsubsidised Private schools*) are free to select students, based on admission exams or any other criterion, and their hiring process is more flexible as it is only ruled by the National Labour Code.

The differences in the management restriction between *Municipal* and *Private schools* might suggest that most-effective schools attractive high quality students also attract the most-effective teachers by offering them higher salaries. Richer Municipalities might also compete for better students and teachers as they can provide additional resources to their schools. Unfortunately, we are unable to observe salaries in the private sector or additional payments to teachers in some Municipal schools; however, this is an issue important issue to be considered throughout the analysis.

The introduction of the voucher programme markedly increased the number of private schools. However, studies of whether quality itself has been enhanced across all types of school as a result of the competition among schools present mixed findings (McEwan and Carnoy (2000); Hsieh and Urquiola (2006)). There is no evidence that definitively shows that Private Voucher schools perform better than Municipal schools. Today, more than 30 years after its implementation, the Chilean national school voucher programme, its benefits, and the potential adverse consequences are all still under discussion.

1.1.2 The demand side

School choice is one of parents main concerns, and despite the increased number of schools they can consider as a result of the introduction of the voucher programme, the decision is still far from straight forward. Traditional Municipal schools and the most prestigious private schools are highly selective, even if Municipal schools are not supposed to be allowed to select pupils. However, the most sought-after

Municipal schools only serve from 7^{th} grade onwards and thus do not affect our analysis of earlier primary grades (e.g. 3^{rd} and 4^{th} grades).

The fundamental decision of which school pupils attend begins in early years when choosing primary schools. This is either because most of the well-known private schools have fewer vacancies in later years, or because traditional Municipal schools select students via admission exams .

Under this framework the perception of school quality plays an important role in school choice. The only publicly available indicator of school quality is the school league tables obtained from the Sistema de Medicin de la Calidad de la Educacin (Education Quality Measurement System known by the acronym, SIMCE), a standardised national examination. Even though the SIMCE is not designed to construct school rankings, SIMCE results nonetheless are usually interpreted to compare schools directly.

1.2 Standardised National examination (Simce)

Since its implementation in early 1980s, students take the Simce exam every year but not in every grade. Initially, students take the exam only in 4^{th}, 8^{th} and 10^{th} grade alternately, although the test has expanded over the years so that students take it more frequently. For example, from 2005 onwards the 4^{th} grade is constantly taken and recently new grades and subjects (e.g. Science and History) were incorporated. From 2012, 2^{nd} grade started being evaluated and currently, since 2013 pupils take the exam in 4^{th}, 6^{th}, 8^{th} and 10^{th} grades. It is important to note that our available information is from 2003 to 2012.

The Simce is administrated by external examiners from the Ministry of Education (Mineduc). The Simce is administrated by external examiners from the Mineduc. The students take the exam in a familiar environment, that is, in their school, classrooms, and with the same classmates. The only differences between the national and school exams might be the formats and presence of an external examiner, who invigilates the National Exam, and who is responsible for the attendance sheet. A minimum pupil attendance on the day of the exam must be met in order for results to be publicly released.[2]

The Simce scores are publicly available at school and grade levels, but not at individual or classroom levels (in the event that a school has more than one class per grade). Both, principals and parents know how well the school is performing

[2]The Simce scores are officially reported if some requirements are satisfied. Among these requirements are: (i) a minimum number of pupils (six per class) must take the exam; (ii) the average school marks of students who are absent on the day of the exam must not be lower than the average of students who attend that day. If there is any irregularity in the Exam process the results are considered invalid and they are not then released.

on average for every grade-cohort that took the exam, but they do not have access to individual scores. The individual Simce scores, to which we have access, are only available for academic purposes, and a formal application must to be submitted to work with them.

Fortunately, all data sets provided by the Mineduc and respective agencies are linked by a unique student, teacher and school by unique identifiers.

1.3 Chilean Education Reform

The Chilean Education Reform Act was introduced in 1981 and is the basis of the current education system. In this section, we describe the Reform and some posterior modifications, especially the most-relevant and latest education policies.

The administration and funding of state schools were decentralised, giving autonomous control to local government. Municipalities thus began to take charge of state schools, from then on called Municipal schools. The administrative mechanism to decentralise public funding to Municipal schools is via vouchers, in the form of a fixed amount of money per enrolled pupil.[3] However, some other sources of supplemental funding from Municipal resources could be used, along with funding from other focused Governmental policies, such as free meal programmes.

The main objective of the Chilean Education Reform was to improve the access to and quality of education by introducing increased competition between state owned school and some of the private schools via individual voucher subsidies. The *Private Voucher schools* were created since then. In the new and open education market all schools have incentives to attract and retain students, in order to assure public or private funds. Parents are, in theory, free to choose schools, as there is no residential restriction to attend schools from different municipalities.

The implementation of the Reform caused a considerable change in the organisation of the education system. As of 1981, three types of schools provide education services in primary and secondary levels. We classify the schools as follows: (1) *Municipal schools:* those schools whose administration was transferred from the Mineduc and depend on the local governments or municipalities. They received the same amount of voucher per pupil enrolled.; (2) *Private Voucher schools:* privately owned schools whose administration is also private (can be religious or secular) but receive fiscal funding from the per-student voucher system

[3]The individual voucher was designed as a specific multiple of a basic unit called Unit of Educational Subsidy (USE). This multiple has to be the same for Municipal and Private Voucher schools, but it could vary according to different types of service (e.g. Primary, Secondary, Technical or Differential Education Schools). The range of its value can go from 50 to 150 per month approximately (Source: Educational Value Subsidy - December 2009. Ministry of Education Chile).

(e.g. the Priority Scholar Subsidy (SEP)); *(3) Unsubsidised Private schools:* privately owned schools that do not receive public funds, and rely on tuition fees charged as their main source of funding.

Examining the evolution in the market composition, we see that the participation of state school in 1981, before the Reform was enacted, it was around 78 percent while the private sector only had 22 percent. After 10 years, in 1991, the distribution of students had changed. Almost 60 percent of students were enrolled in *Municipal schools*; 32 percent were in the *Private Voucher schools*; and 8 percent were in the *Unsubsidised Private schools*.[4] By 2007, the distribution had again changed: the number of students enrolled in *Municipal schools* had decreased to 46 percent of total enrolment, *Private Voucher schools* participation had grown to 45 percent, and enrolment in *Unsubsidised Private* schools had remained unchanged with at 9 percent. The majority of years for which data are available, suggest that the distribution appears to have stabilised, with only small changes from *Municipal* to *Private Voucher schools* affecting the distribution.[5]

1.3.1 Modifications post Reform

Since the reform was implemented in the early 1980s, some modifications have been made, especially during the 1990s. Even though the competitive education system has been retained, the government increased the amount of resources spent in education on less favoured pupils, developing both universal and targeted programmes.

Some of the new policies implied different sources of funding. In 1993 the Shared Funding System (SFS) was introduced, and Private Voucher schools were free to join. The SFS allows schools to charge parents extra tuition fees without losing the total value of the voucher. Participant schools would reduce the amount of money received per voucher, depending on the additional amount of fees charged. The maximum fee allowed is around four times the value of a voucher, and under the co-payment tier the voucher would be decreased up by to its 40 percent. The introduction of the SFS could be considered very successful, as almost 40 percent of the Private Voucher schools converted to the new system by 1996.[6]

Simultaneously, the *Municipal schools* also get financial assistance from municipalities and from the Mineduc given targeted programmes, mainly designed for the most vulnerable students. Those specific programmes increase the real edu-

[4]From the share evolution graph presented in McEwan et al. (2008).

[5]Refer to Chapter 3. Data description. In Table 3.7 we show the distribution of students by type of school dependence from 2003 to 2012 using our available dataset.

[6]See Sapelli and Vial (2002)

cation expenses and reward those institutions with higher academic achievements within a comparable contextual group. Therefore, *Municipal school* budgets differ among and within municipalities.

Municipal schools cannot select students, and they can reject applicants only when their maximum enrolment capacity has been reached (and it has to be proved). In contrast, *Private Voucher schools* are able to select students by admission exams, interviews, enrolment fees and any other criterion. The pupil selection also occurs when approximately 67% of *Private Voucher schools* charge an additional fee under the SFS scheme.

Teachers from Municipal schools are represented by a centralised union which negotiates salaries and benefits for them. Whereas teachers from Private Voucher schools are hired under regular private contracts, providing more-flexible wages and benefits. As our research interest is in teacher and school effects, it might be interesting to study whether different teacher employment contracts could have an impact on teacher effectiveness on pupil academic achievement, although we do not have this information.

1.3.2 The latest educational policies

The Priority Scholar Subsidy (PSS)

The Priority Scholar Subsidy (PSS) Law, enacted in February 2008, offers additional economic resources to schools with the most socioeconomically disadvantaged students. The extra funding is assigned to improve educational quality in schools where it is expected to be most costly to teach and generate students achievement gains. School participation in this programme is voluntary, but once the school decides to join it has to fulfil certain legal requirements.[7]

Because the programme consists of increasing the amount of the voucher per priority student, extra funding depends on the number of priority students enrolled in a PSS-School. The priority classification is obtained via specific criteria which is annually revised. *Municipal* and *Private Voucher* schools' participation depends on the number of priority students enrolled in their schools, and their willingness to follow the administrative recommendations instructed by the Ministry of Education.[8]

[7]Referred to Priority Scholar Subsidy (PSS) Law, from the document LEY DE SUBVENCION ESCOLAR PREFERENCIAL (SEP). LEY No. 20,248, Febrero 2008.

[8]Sequential criterion for categorising Priority Student Rights enrolled in SEP-Schools: (i) They do not pay fees when the school is subject to the Shared Financing System; (ii) Priority Students cannot be expelled, and if they repeat a grade, they should be retained in the school; (iii) The School commits to improve their academic results; (iv) The School gets extra funding to elaborate and implement an Improvement Educational Plan.

The SEP-School commit to: (1) Present and execute an Improvement Education Plan, elab-

PSS-Schools have to show improvement in the academic performance of priority students. So, the expectation of succeeding in this challenge determines their participation. The additional amount of subsidy per student depends on the PSS-School category, and it could mean an increase in the amount of the individual voucher up to 56%. A second criterion is related to the concentration of priority students within school. Schools in which priority students represent at least 60% of the total enrolment could apply for an extra 30% increase in the voucher.

Extensions for the Standardised National Exam (Simce)

- From 2010, the Simce for Sports and Physical Activities is taken by students in 8^{th} grade in order to evaluate the physical condition of students, and to document differences across schools.
- From 2011, the Information and Communications Technology (ICT) Simce Exam started to assess the different ability levels related to ICT. Students take this exam in in 10^{th} grade.
- From 2012, a new Reading Simce Exam is given to students in 2^{nd} grade in an effort to enable early identification of any reading difficulties.

Additional Supporting Programme

In 2011 a voluntary supporting programme was created for low achievement *Municipal* and *Private Voucher schools.* The Share Support Plan (SSP) (Plan de Apoyo Compartido - PAC) offers principals the opportunity to work jointly with a specialised team from the Mineduc to design and develop strategies that improve the Simce scores in 4th grade. The target is students from Kindergarten to 4^{th} grade in low achievement schools. Currently, more than 1,000 schools participate, representing approximately 210,000 students and 6,000 teachers.

Compulsory increase in teaching Language hours

From 2012, there is a mandatory increase of 240 hours in Language (yearly) instruction from 5^{th} to 10^{th} grades, and 320 hours in Maths instruction in the same

orated mainly by the principal; (2) Establish academic attainment goals, particularly to those classified as priority students; (3) Inform to the community about the programme; (4) publicly present a yearly report of the SEP resource usage; (5) Formalise the activities of School Council, Teachers Council, and Parents Centre; (6) Incorporate artistic activities besides and/or cultural and sports activities; (7) Incorporate artistic activities besides and/or cultural and sports activities; (8) Resign to select pupils in admission process.

Source: SEP Brochure, prepared by Ministry of Education Chile, November 2011. *"Folleto Informativo SEP para Padres y Apoderados"*.

grades as well.

Compulsory nursery school

From 2014 Pre-Kindergarten and Kindergarten education should be provided in every school, ensuring universal and free access from age 3. Kindergarten is now compulsory, extending the total compulsory education in Chile from 12 years to 13 years.

Chapter 2

Teacher Effects and Value-Added Modelling: A review of specifications and estimators

Abstract

In this chapter we present a detailed review of "Teacher Effects" (TEs) estimations based on common Value Added Models (VAMs) found in the literature. From a general cumulative education production function, we derive a general achievement function (GAF), and show how the four most common VAMs can be derived by imposing various restrictions.

We then discuss three different approaches to estimate TEs in these models: **Approach 1**, which relies only on observable teacher characteristics; **Approach 2**, which uses teacher dummy variables to identify teacher effects; **Approach 3**, which assumes TEs are randomly drawn from a particular distribution. Then, we review the literature based on the last two approaches, which are the most common approaches.

The third part of this review is based on the validation of TEs estimates and the predictability of teacher rankings, depending on the educational context. The VAMs have to be validated because every study has its own context, and it is important to assess how trustworthy TE predictions are under different contextual environments. Most of the validation studies conclude that under random assignment of teachers to classroom there are not considerable differences between estimators, once one controls for enough student, teacher and school characteristics (including previous pupil attainment).

Contents

2.1 Introduction

Human capital acquisition is considered as a dynamic cumulative process, which is commonly expressed by an education production function where student achievement is explained by pupil, teacher and school characteristics (e.g. Hanushek (1971); Todd and Wolpin (2003); Sass et al. (2014)).

Earlier studies have suggested that conventional concepts of school and teacher quality, such as teacher qualifications and years of experience, are weak indicators of true "School Effects" (SEs) and "Teacher Effects" (TEs) on pupils' academic outcomes. For example, Hanushek (1986) recommends that TEs must be estimated within a Value Added framework that truly identifies the aggregated value of each teacher on improving average student performance, compared to other teachers in a similar context (controlling for school, classroom and pupil characteristics).

Based on Value Added Models (VAMs), more recent investigations (Rivkin et al. (2005); Aaronson et al. (2007); Buddin and Zamarro (2009)) have proposed that observed teacher characteristics tend to have low predictive power in forecasting student achievement. The VAMs rely on the concept that TEs must represent individual teacher skills rather than student average improvements due to pupil sorting and teacher-to-classroom assignments.

In particular, Rivkin et al. (2005) suggest that most of the teacher contribution to student performance comes from teacher unobserved heterogeneity. There is not much evidence that teacher observable characteristics have an effect on pupil academic achievement.[1]

Students outcomes might be affected by factors other than teacher and pupil skills. A vast literature suggests that unobserved student heterogeneities are

[1] Clotfelter et al. (2006) is one of the few studies which found significant impact of teacher observable characteristics on student academic performance.

also related to *non-cognitive* abilities (e.g. Heckman et al. (2006)). These abilities are also influenced by social skills which are mainly developed within schools. Hence, the combination of student, teacher and school unobserved characteristics is expected to have significant impact on academic outputs. However, we only focus on measuring impacts on individual *cognitive* abilities.

Within the VAM context, we identify three levels of unobserved heterogeneities affecting pupils educational progress. Student and teacher latent abilities joined in a particular school environment contribute to pupil attainment progress. Thus, in addition to observable characteristics, it is important to consider all of these potential unobserved factors when modelling pupil academic performance.

The recent availability of richer data sources, in particular from administrative datasets, and better access to computational technologies have generated a considerable increase in education studies. The technical scope is larger, and now it is possible to estimate TE distributions based on computationally intensive methods using iterative algorithms. These algorithms allow us to predict teacher impact, and to construct teacher rankings according to their expected effectiveness.

In this chapter we present a detailed review of TE estimators based on popular VAMs. We discuss the most used VAM specifications and estimation methodologies. We propose a set of restrictions generally imposed to VAMs and we discuss what type of assumptions authors have claimed in their VAM estimation strategies.

Basically, we identify three main approaches to VAM estimators: **Approach 1** which is based on observed teacher characteristics only; **Approach 2** which is supported by introducing teacher dummies which capture their corresponding teacher effects; and **Approach 3** which relies on the assumption that TEs are randomly drawn from a stochastic distribution.

Then, we show how authors have tried to validate teacher Value Added measures and teacher rankings based on TE estimates. This analysis is highly relevant as every study has its own context, and it is necessary to determine how reliable TE predictions are under contextual environments.

After reviewing most of the current discussion on TE estimations and their statistical validations, we find that there is not a unique consistent estimation strategy that can be employed across all contexts. We also find that it is not advisable to discard VAM specifications and estimators using typical validation tests when potential non-random assignment of teachers to classrooms occurs. Therefore, it is still an open discussion of how helpful and reliable TE estimates are under different scenarios.

The purpose of this Chapter is to provide an updated review of the teacher effectiveness literature, and to introduce a general VAM which is also going to be employed to estimate our own TEs and SEs in later chapters.

The current chapter is organised as follows; In Section 2 we introduce a general achievement function from which can be derived most of the VAM specifications used in the literature. Section 3 explains the teacher effectiveness concept within the VAM and presents the most common estimation methodologies. In Section 4 we review VAM estimators classified by different estimation approaches, and we check the discussion of VAM validation. In Section 5 we present some of the Value Added measures obtained in the literature which can be compared to our TE estimations presented in Chapter 5. Finally, Section 6 provides conclusions and further remarks.

2.2 Value Added Models

Our research interest is focused on "Teacher Effectiveness" and the corresponding TE measures which are usually estimated within a Value Added Model (VAM) framework. We therefore start with the most commonly estimated VAMs and discuss how these can be derived by imposing some structural restrictions on the General Achievement Function. The estimation of TEs will be discussed in the next Section.

The theoretical model that forms the basis for the Value Added Models (VAMs) is the one initially proposed by Hanushek (1971), and it relates educational outputs (achievements), such as Exam Scores, to a set of inputs, current and historical, provided by family and school characteristics. The model also includes a component of individual ability, though it is ignored for estimation purposes. Boardman and Murnane (1979) extended the general achievement model to include unobserved characteristics and discussed the estimation and interpretation of the effects of inputs in the presence of different types of data such as a simple cross section or a panel. Other empirical papers following this model are: Todd and Wolpin (2003); Hanushek and Rivkin (2010); Harris and Sass (2014). These more recent papers also base their VAMs on similar production functions, assuming that the current student achievement is a function not only of current educational inputs but also of all past inputs. This lead to the notion of a "Cumulative Production Function" (CPF).

The CPF specifies the achievement $A_{i,g}$ for child i, who is in grade (class or year) g, as:

$$A_{i,g} = F_g(\mathbf{X}_i(g), E_i(g), \alpha_i, \varepsilon^*_{i,g}) \tag{2.1}$$

where the vector $\mathbf{X}_i(g)$ includes all current and past individual and family educational inputs. Similarly, the vector $E_i(g)$ includes the entire history of teacher and school inputs. The input factor α_i is assumed to be a time-invariant student specific unobservable variable (e.g. student genetic endowment), and $\varepsilon^*_{i,g}$ can be interpreted as a measurement error or a component of luck (e.g. an individual's bad day or a disruptive environment at the moment of an exam).

Assuming a linear separable function for $A_{i,g}$, we transform equation (2.1) into our General Achievement Function (GAF).

$$A_{i,g} = \sum_{r=1}^{g}[x'_{i,r}\beta_{g,r} + e'_{i,r}\gamma_{g,r}] + \alpha_i + \varepsilon_{i,g} \tag{2.2}$$

The vector $x_{i,r}$ contains the individual and family educational inputs, and $e_{i,r}$ consist of teacher and school inputs in grade r. The coefficients $\beta_{g,r}$ and $\gamma_{g,r}$ are vectors representing the impact of each observed variable, in all current and past grades, on current academic performance. The observable and unobservable child/family specific components from $\mathbf{X}_i(g)$ have been separated out and we have assumed that there is only one unobserved child specific component α_i and no family-specific unobservable component.

We follow the literature and assume that the effect of α_i is grade invariant. The term $\varepsilon_{i,g}$ is assumed to be an idiosyncratic *iid* (independent and identically distributed) error in the approximation. It is important to highlight the assumption that current inputs not only have an effect on current achievement, but also continue to have an impact on future achievements too, *ceteris paribus*.

Although equation (2.2) does not explicitly show it, the school educational input vector $e_{i,r}$ includes observable as well as unobservable variables. For example, these unobservables which are present in two levels, may include: (i) at the school level characteristics, the ability of the principal, and the availability of resources; (ii) at class-level variables, teacher ability, and peer characteristics in the classroom.

The estimation of equation (2.2) requires a very rich data source, and so far, no one has been able to access this type of data with all historic records, and even if it were possible, estimation would be computationally challenging. Researchers have had to impose further restrictions on the parameters to make it estimable. The important issue that needs to be addressed is whether data limitations and the restrictions imposed prior to estimation will deliver an estimator that is consistent for the parameters of interest. A systematic discussion of these issues is presented in the next subsections.

2.2.1 Restricted VAMs

We take the GAF from equation (2.2) as a starting point to first discuss the set of restrictions generally imposed to obtain a manageable Value Added Model for estimation purposes. Then we discuss specific restrictions on VAMs that have been imposed in the literature due to data limitations.

The first restriction imposed is that although there are grade-specific effects of inputs, the persistence of effects on subsequent pupil achievements decays over time. In addition, it is also assumed that all input factors have the *same* rate of persistence parameter λ across periods, and it follows a geometric distribution. This particular form of decaying has the considerable advantage of enabling us to rewrite the initial GAF from equation (2.2) with two main components, all current inputs and a function of the previous years' achievements.

Under the above assumptions, equation (2.2) can now be written as

$$A_{i,g} = x'_{i,g}\beta_g + e'_{i,g}\gamma_g + \alpha_i + \sum_{r=1}^{g-1} \lambda^{g-r}[x'_{i,r}\beta_r + e'_{i,r}\gamma_r] + \varepsilon_{i,g} \tag{2.3}$$

where $0 \leq \lambda \leq 1$ and $\beta_{g,} = \lambda^{g-r}\beta_r$.

Equation (2.3) says, for example, that the effect of last year's inputs $x_{i,g-1}$ on $A_{i,g}$ is $\lambda\beta_{g-1}$ and the effect of $x_{i,g-2}$ on $A_{i,g}$ is $\lambda^2\beta_{g-2}$. This assumption has two important implications: (i) the rate of persistence is the same for all educational input factors, and (ii) the rate of geometric decay is constant across grades: hence, the effect of an input factor in, for example, two previous periods will have the same effect $\lambda^2\beta_{g-2}$ on current grade g achievement, independent of whether the current grade is 3, 4 or 5.

Thus, considering equation (2.3) for period *g-1* multiplied by λ to subtract both sides of equation (2.3), we get[2]

$$A_{i,g} - \lambda A_{i,g-1} = x'_{i,g}\beta_g + e'_{i,g}\gamma_g + (1-\lambda)\alpha_i + \varepsilon_{i,g} - \lambda\varepsilon_{i,g-1}$$

and we rewrite the above equation as:[3]

Model 1:

$$A_{i,g} = \lambda A_{i,g-1} + x'_{i,g}\beta_g + e'_{i,g}\gamma_g + \alpha_i + \varepsilon_{i,g} - \lambda\varepsilon_{i,g-1} \tag{2.4}$$

Equation (2.4) now forms the basis for various VA estimations. This equation represent the most general case of VAMs, which can be interpreted as indi-

[2]This trick is used in the Partial Adjustment Models to reduce the number of parameters, and the demonstration of the transformation is shown in Appendix 2.2.

[3]WLOG we can write $(1-\lambda)\alpha_i = \alpha_i$. However, as we shall see later, this term will disappear when λ is set equal to 1 (see **Model 3**).

cating that: pupil achievement in a given grade g depends on current and previous observed and unobserved input factors (including teacher effects).

Before we turn to the estimation of TE, it should be noted that the estimation of equation (2.4) is complicated due to the presence of $A_{i,g-1}$ and α_i, which are correlated with $\varepsilon_{i,g-1}$. Hence, it requires much more stringent assumptions to obtain consistent estimators, as we will discuss later in the VAM specifications and literature review (e.g. Aaronson et al. (2007); Kane and Staiger (2008); Rothstein (2010); Buddin (2011)).

We next examine what additional restrictions researchers have imposed on equation (2.4) to enable them to address this issue. Of course, if a valid instrument for $A_{i,g-1}$ can be found, estimation of this equation becomes easier. In that sense, we present some of the restricted models commonly used as VAMs in the literature.

Model 2: Non-Historical with total decay of input factor effects ($\lambda = 0$)

$$A_{i,g} = x'_{i,g}\beta_g + e'_{i,g}\gamma_g + \alpha_i + \varepsilon_{i,g} \tag{2.5}$$

Model 3: No decay with full persistence of input factor effects ($\lambda = 1$)

$$A_{i,g} - A_{i,g-1} = x'_{i,g}\beta_g + e'_{i,g}\gamma_g + \varepsilon_{i,g} - \varepsilon_{i,g-1} \tag{2.6}$$

Model 4: Geometric decay with a fixed known persistence parameter λ_0, $(0 < \lambda_0 < 1)$

$$A_{i,g} - \lambda_0 A_{i,g-1} = x'_{i,g}\beta_g + e'_{i,g}\gamma_g + \alpha_i + \varepsilon_{i,g} - \lambda_0\varepsilon_{i,g-1} \tag{2.7}$$

In **Model 4**, λ is set equal to a specific parameter value to avoid the correlation problems mentioned above (e.g. λ_0 is taken from estimates reported in the literature).

However, further additional assumptions for the above models are necessary in order to generate consistent estimators depending on the estimation approach taken. In the following section we explore in detail the estimation methodology of TE and the required assumption for each estimation approach.

2.3 Teacher Effectiveness in VAMs

So far, we have included $e_{i,g}$ in the VAMs to capture the effect of school and teacher inputs that can be composed of both observable and unobservable variables. Suppressing the school specific index, we let S_g and $T_{j,g}$ be the observed, and s_g and $\tau_{j,g}$ the unobserved school and teacher characteristics, respectively. The sub-index

j identifies the teacher, and g the grade in which student i is enrolled in. We now write teacher and school inputs factors as

$$e'_{i,g}\gamma_g = T'_{j,g}\pi_g + \tau_{j,g} + S'_g\theta_g + s_g \tag{2.8}$$

The total contribution of teacher j in this model is given by $T'_{j,g}\pi_g + \tau_{j,g}$, whereas the total contribution of the school in grade g over and above individual teacher effects, is represented by $S'_g\theta_g + s_g$.[4] How to disentangle both contributions and, in particular, how to estimate TEs has been a highly debated research question in the recent literature (e.g. do we measure teacher effectiveness as $T'_{j,g}\pi_g + \tau_{j,g}$ conditional on school characteristics or should we use another alternative specification?).

The idea is that the potential effectiveness of schools and teachers should not depend on the type of students enrolled in a particular type of school, and it should not depend either on teacher allocation into sorted groups of students within schools. True teacher effects should be independent of the contextual and environmental setting. Only when this occurs we can talk about causal effects rather than correlations.

For the moment, and just to keep the discussions focussed on TE estimations, we will ignore the additional contribution of schools ($S'_g\theta_g + s_g$) in the following subsections.[5]

The most popular methods of estimation have been either Ordinary Least Squares (OLS) or Maximum Likelihood Estimation (MLE). Whatever is used, we have to ensure that if unobservable components are present in the estimation equation, they must be uncorrelated with other included covariates in order to obtain consistent estimators.

One possible reason for the failure of this assumption is that teachers are not always randomly assigned to classes. If this is the case, even in the absence of any other correlation, the unobserved child specific component α_i in equation (2.4) will be correlated with the observed teacher and pupil characteristics, including past achievements. We discuss this important issue in detail below.[6]

We next turn to the estimation of TEs using the VAM and discuss the

[4]Note that we associate teacher j with class j, and there can be more than one class or teacher j per grade g in the same school.

[5]Similar arguments in terms of the econometric issues arise in the presence of school observables and unobservables in these VAMs. However, if teachers and students are randomly allocated to schools, we don't have to worry about the school effects.

[6]It may be possible to include many student and teacher characteristics in equation (2.4) to capture most of the influence of unobserved α_i in the model (Guarino et al. (2014b)). However, the issue of correlation between past achievement and the lagged equation error is still present when a rich set of variables is included in the models.

importance of further assumptions that are needed to obtain consistent estimators.

2.3.1 Estimation of VAMs and TE

Models 1-4 can now be rewritten as

Model 1 (**unrestricted** λ):

$$A_{i,g} = \lambda A_{i,g-1} + x'_{i,g}\beta_g + T'_{j,g}\pi_g + \tau_{j,g} + \alpha_i + \varepsilon_{i,g} - \lambda\varepsilon_{i,g-1} \tag{2.9}$$

Model 2 ($\lambda = 0$)

$$A_{i,g} = x'_{i,g}\beta_g + T'_{j,g}\pi_g + \tau_{j,g} + \alpha_i + \varepsilon_{i,g} \tag{2.10}$$

Model 3 ($\lambda = 1$):

$$A_{i,g} - A_{i,g-1} = x'_{i,g}\beta_g + T'_{j,g}\pi_g + \tau_{j,g} + \varepsilon_{i,g} - \varepsilon_{i,g-1} \tag{2.11}$$

Model 4 ($\lambda = \lambda_0$, where λ_0 is a **preset value**):

$$A_{i,g} - \lambda_0 A_{i,g-1} = x'_{i,g}\beta_g + T'_{j,g}\pi_g + \tau_{j,g} + \alpha_i + \varepsilon_{i,g} - \lambda_0\varepsilon_{i,g-1} \tag{2.12}$$

We first list the main common assumptions that are required for all approaches to modelling TEs, and then we discuss the specific assumptions required for each specific approach.[7]

Assumption A1. Strict Exogeneity of covariates conditional on unobservables:

$$E[\varepsilon_{i,g}|past,\ present \text{ and } future \text{ values of } (x, T, S), \alpha_i, \tau_j, s_g] = 0.$$

This assumption implies that all covariates are strictly exogenous conditional on all the time invariant unobservables. This statement is about the conditional mean of the time-varying idiosyncratic error term $\varepsilon_{i,g}$. In certain cases, it might be possible to relax this strict exogeneity to sequential exogeneity whereby the conditioning is only carried out in terms of current and past values of the covariates and not in terms of future values. Hence, when the lagged dependent variable, such as $A_{i,g-1}$, is included in the VAM, it satisfies sequential exogeneity rather than strict exogeneity.

Also note that the variables relevant to other students are assumed not

[7] Note, we have already assumed $\varepsilon_{i,g}$ to be *iid*.

affected by the outcome of student i.

Assumption A2. Random assignment of students to teachers (or classrooms)

This assumption states that all student variables are uncorrelated with all teacher variables. It is important to highlight that just having random allocation of teachers to classes may not be enough to yield consistent estimators of parameters of interest in the above models. We need to ensure that all unobservables factors in the estimation equation are uncorrelated with all observable covariates included in the model, in order to obtain consistent estimators.

We now turn to the discussion of three general approaches that have been used for modelling and estimating TE, and we discuss the additional specific assumptions that are needed to obtain consistent estimators in each case.

Approach 1

Early investigations were focussed on identifying high quality teachers using observable characteristics only (Hanushek (1986)). This approach assumes that a very rich data source is available to account for the *full effects* of teachers on achievement using a variety of teacher characteristics, i.e. we can ignore $\tau_{j,g}$ in the above models.

Under *Assumption A1*, OLS estimation of **Model 3** either with a pure cross-section that has one single observation on the previous score ($A_{i,g-1}$) or with panel data, will produce consistent estimators. We can also use OLS to estimate **Model 2** and **4**. However, if we only have a single cross-section with the previous grade achievement, we will need the additional assumption ***A3.1*** due to the presence of α_i.

Assumption A3.1. Expected value of student heterogeneities is equal to 0, conditional on observables:

$E[\alpha_i|past, present$ and $future$ values of $x, T] = 0$.

Note that we have omitted τ_j from the conditioning variables for this particular approach. If the model estimated contains the student unobservable factor α_i, in case it cannot be eliminated prior to the estimation, then ***A3.1*** becomes crucial. This is because α_i might still be correlated with student covariates even if we assume that teacher and student variables are uncorrelated because of *ran-*

dom allocation of student to teachers as stated in ***A2***, although conditioning in T would not be needed.

On the other hand, ***A3.1*** is not requiered if we have multiple observations on achievement for each student, as we can use within-group transformation to eliminate α_i prior to the OLS estimation. Also note that conditioning on T would not be needed either if *students and teachers are randomly assigned to each other* as ***A2*** states.

Due to the presence of lagged achievement $A_{i,g-1}$ and $\varepsilon_{i,g-1}$ (see equation 2.9) we cannot estimate **Model 1** parameters with OLS. However, an instrumental variable (IV) estimation can be used if a suitable instrument is proposed for $A_{i,g-1}$.

Recent results suggest that traditional measures of teacher quality, such as additional teacher qualifications and years of experience, do not generally make a significant contribution to improving student performance. Therefore, it would be crucial to allow for unobservable teacher characteristics in these models (Rivkin et al. (2005); Aaronson et al. (2007); Buddin and Zamarro (2009)).

Approach 2

This approach replaces the observable and unobservable variables by a set of binary indicators for teachers, the so-called "Fixed Effects" (FE) approach. The main advantage of this approach is that we do not need to worry about possible correlation between unobserved (or unaccounted for) teacher characteristics and the rest of the variables included in the model because teacher indicators will control for these latent variables.

VAMs can now be estimated by OLS (also known as the Within-Group Estimator (WG)). The same estimator for TE can also be obtained using the following equivalent procedure where we first obtain the residuals from an OLS regression of teacher dummies on all of the observed covariates and then regress the test scores on these residuals.[8] This estimation requires multiple observations

[8]This result is due to Frisch-Waugh Theorem which shows the following:

Consider the model $y = X_1\beta_1 + X_2\beta_2 + u$. Then the OLS estimator of β_1 is: $\hat{\beta}_1 = (X_1'M_2X_1)^{-1}(X_1'M_2y)$; where $M_2 = I - X_2(X_2'X_2)^{-1}X_2'$ which is the symmetric idempotent (projection) matrix, i.e. $M_2M_2 = M_2$ and, M_2X_1 is the residual from the regression of X_1on X_2. It is therefore clear that the above is also equivalent to a two-step procedure where we carry out a regression of X_1 and y, on X_2 to obtain the residuals M_2X_1 and M_2y in the first step and then regress the second set of residuals (M_2y) on the first set of residuals (M_2X_1) to obtain the same estimator. Because of the property of M_2, we can also obtain the same estimator by regressing y on M_2X_1. In our example X_1 is the set of "teacher dummies".

Note that estimating $y = X_2\beta_2 + v$ by OLS and then obtaining the residuals (M_2y) in step one and then regressing this on X_1 to obtain an estimate of (β_1) will not give the right $\hat{\beta}_1$ since the new $\tilde{\beta}_1$ would be $(X_1'X_1)^{-1}(X_1'M_2y)$.This latter procedure is sometimes used for the estimation of TE and is known as Average Residual (AR) estimator. For example, the first stage residuals will be regressed on a set of teacher dummies which is equivalent to taking the teacher averages of the residuals (See Guarino et al. (2014b) for a discussion of this and other related points).

for each teacher if we want to control for SEs too.

The OLS estimator of β_g in **Model 3** under ***A1*** will be consistent. However, in order to use OLS to estimate β_g in **Model 2** and **4**, we will also require ***A2*** and ***A3.1*** to hold. Note, OLS will give consistent estimator of the effects of covariates but not of the TEs since the estimator of TEs suffers from the *"Incidental Parameters Problem"* (Neyman and Scott (1948)).

The consistency property is considered with respect to the number of students per teacher going to infinity, which cannot happen. The OLS estimation of **Model 1**, even under assumptions ***A1-A3.1*** will not provide consistent estimators due to the correlation between $A_{i,g-1}$ and $\varepsilon_{i,g-1}$. Nevertheless, the estimator of TE would be unbiased.

Approach 3

Here the model specification explicitly accounts for the unobserved teacher effect $\tau_{j,g}$ and assumes this to be randomly drawn, with either including or excluding observable teacher characteristics. This estimation methodology is typically known as the "Random Effects" (RE) approach because teacher effects are treated as random effects. Common estimators for this approach are the Feasible Generalised Least Square (FGLS) and the Maximum Likelihood Estimator (MLE). Additional assumptions are now needed to obtain consistent estimators of the above models.

Due to the presence of $\tau_{j,g}$, we now require some assumptions regarding the relationship between this and the rest of the variables in the equation model to obtain consistent estimators. Then, a new version of ***A3.1*** must hold under this approach.

Assumption A3.2. Expected value of student heterogeneities is equal to 0, conditional on observables and unobservables:

$E[\alpha_i | past, present \text{ and } future \text{ values of } x] = 0.$

If ***A2*** holds, we do not need to condition on T or τ in ***A3.2***. Nevertheless, for estimators following this approach we additionally need to impose ***A4***.

Assumption A4. Expected value of teacher effects is equal to 0, conditional on observables and unobservables:

$E[\tau_j | past, present \text{ and } future \text{ values of } T] = 0.$

If ***A2*** holds, we do not need to condition on x and α in ***A4*** .

The common FGLS and MLE estimators rely on the assumption that teacher and student unobserved heterogeneities are drawn from distributions with homoskedastic variances and independent of each other.[9] It is also customary to assume these unknown distributions to be *Normal* for the use of maximum likelihood methods.

As before, **Models 2**, **3** and **4** can be estimated using FGLS or MLE. Under assumptions ***A1-A4*** (and *Normality* assumption for MLE) the estimators will be consistent. If model assumptions hold, this approach will provide more efficient estimators relative to the **Approach 2** (Maddala et al. (1997); Guarino et al. (2014a)).

In summary, one can either include teacher binary indicators (the so-called FE approach) or treat the teacher unobservable characteristics as random effects (RE approach) with or without including teacher observable characteristics. The RE approach using standard techniques (FGLS or MLE) will not yield consistent estimators if any of the random unobserved components is correlated with included variables in the model.

2.3.2 Teacher effects (TE): One-Step vs Two-Steps estimations

If the FE methodology (**Approach 2**) is used, the TE would be predicted using the estimated effects of the binary indicators. On the other hand, if the RE methodology (**Approach 3**) is followed estimating the VAM model using FGLS, a best linear unbiased prediction (BLUP) of the TE is made. The BLUP estimator in the linear model with teacher random effects, is the Empirical Bayes (EB) estimator of the TE which is also known as the Bayesian shrinkage estimator which we discuss in the following subsection.

This estimator shrinks the initial estimator towards the overall mean, and that is the reason it is called Two-Step estimation of TE.[10] The adjustment made depends on how noisy the TE estimates are. For example, when the number of students per teacher is small and her individual error variance is too high, the TE measure becomes less informative, then her TE estimates is shrunk towards the overall mean. The shrinkage estimators can also be calculated when the model is

[9] Even if the unobserved teacher heterogeneity is assumed to be a random draw, one can still estimate the model using Within-Group estimation including teacher dummies. TE estimators will be unbiased but not efficient as the best linear unbiased estimator ($BLUE$) is the GLS estimator.

[10] The main reason given for this is that the reliability of TE estimates may not be good due to relatively small number of observations per teacher, which by construction cannot asymptotically tend to infinity.

estimated by MLE, in which case this is the known as the MLE-EB estimator and it is equivalent to the expected mean of the posterior distribution of τ given the data and the Maximum Likelihood parameter estimates.

However, note that the model assumptions to generate a consistent estimator using FGLS or MLE methods need to be satisfied. In particular, the failure of non-random assignment of students to teachers might make these procedures invalid.[11] We discuss these issues in the next subsection.

It is important to highlight that the weight of the shrinkage factor not only depends on the number of observations per teacher, but also depends on how informative the total variance of teacher effects is, with respect to the total variance of the model. If the variance of the teacher effects is relatively high compared to the total error variance, it becomes less informative. Hence, the shrinkage factor increases, making the TE estimates closer to those obtained with One-Step estimations (or unadjusted estimates).

Within Group (WG) estimators (**Approach 2)** and Average Residual (AR) estimators (see footnote 9) can also be shrunk, and hence called Two-Step estimations of TEs. For example, TE estimates obtained from an AR estimator are shrunk from their original estimation based on model variance. Shrunken TE estimates obtained from initial AR estimators are called SAR or AR-EB. Similar to MLE-EB, the AR-EB is also known as a type of empirical Bayes estimation.

It has been shown that the results obtained by these Two-Step estimations are similar to the the One-Step Iterative Bayesian procedure, and both are preferred to the classical estimation or One-Step approach (OLS, MLE) without adjustment to their initial estimates (Maddala et al. (1997)). Within the class of Two-Step estimations, if there is non-random assignment of teacher to classroom then the TE estimates obtained under **Approach 3** are less biased than those obtained by **Approach 1** (Guarino et al. (2014a)).

An obvious concern here is whether the assumptions (e.g. the particular distributional assumptions used in the MLE and the orthogonality of errors to covariates) that are needed to obtain a consistent estimator when using likelihood techniques are valid or not.

As our later estimations in Chapter 5 and 6 will be focused on the MLE-EB estimations, in the following subsection we explain the empirical Bayes estimation. Here we try to understand under which context the empirical Bayes estimation was developed, and we describe its estimation process with a simplified version of the Value-Added model.

[11] When the number of students per teacher tends to infinity, the BLUP/EB estimator will tend to the FE estimator of the TE (Robinson (1991)).

2.3.3 Empirical Bayes estimation

The empirical Bayes estimation of teacher effectiveness (TE) has become very popular within the VAM literature and other related topics, although the concept of predicting random effects have been discussed since early 1980s (Morris (1983); Robinson (1991))

As the data are usually organised in clusters, the effects of these groups on a particular subject could be estimated by two alternatives approaches: (a) fixed effects, and (b) random effects. Whether to treat clusters as fixed effects or random effects depends on the question we want to answer with our model, then a suitable estimation methodology must be defined.

In those cases where the cluster or unobserved heterogeneity effects are not our main interest, and we want to control for potential bias when they are correlated to other explanatory variables, is recommendable to treat them as fixed effects. The common estimators in these cases are the Ordinary Least Square (OLS) with a dummy identification variable per group, and the Within-Group (WG) estimator which estimates mean deviations within groups in a panel data framework.

Alternatively, if we are interested in making inference on clusters effects and its dispersion is relevant for our analysis, we should treat the unobserved heterogeneity as random effects. Prediction of random effects can be made from different perspectives (e.g. Bayesian methods, empirical Bayesian, and frequentist prediction/estimation).

In pure Bayesian framework, all covariates are considered random, and objective parameters (mean and variance) of prior distributions are required. Then, multiplying the prior densities with the maximum likelihood estimation we derive the posterior distribution. All inference regarding predicted random effects is made from the posterior distribution of the Bayesian estimation.

On the other hand, empirical Bayes estimation considers the prior distribution as unknown, and uses the available data to estimate it. Therefore, empirical Bayes approach involve Bayesian concepts in a frequentist framework, where the estimation of random effects requires the recovery of the distribution of estimated parameters.

The linear regression models which combine fixed effects and multivariate random effects are known as: Bayesian hierarchical models, Mixed models, Linear random-intercept models, and Multilevel generalised models, among others. However, all of them are Random Effects models that can be estimated by the same methodologies mentioned above.

Empirical Bayes estimation in a simplified framework

Setting a simple case for a linear mixed model to predict random teacher effects τ_j.

$$y_{i,j} = x'_{i,j}\beta + \tau_j + \varepsilon_{i,j} \tag{2.13}$$

$x_{i,j}$ corresponds to observable covariates with β fixed effects for student i, taught by teacher j. Where $i = 1...N$ and $j = 1...J$. In a matrix representation we have

$$\mathbf{y} = \mathbf{x}\beta + \Gamma + \epsilon \tag{2.14}$$

In the Bayesian approach the predictor for τ_j, will be given by the posterior density function:

$$f(\tau_j|\mathbf{y}, \mathbf{x}; \beta) = \frac{f(\mathbf{y}|\tau_j, \mathbf{x}; \beta)f(\tau_j)}{\int f(\mathbf{y}|\tau_j, \mathbf{x}; \beta)f(\tau_j)\,\mathrm{d}\tau_j} \tag{2.15}$$

Following Bayes Theorem and assuming all densities are proper probability densities functions, such us $f(\tau_j) > 0$ and $\int f(\tau_j)d\tau_j = 1$, we have that the posterior distribution of τ_j conditional on the data is proportional to the conditional likelihood times the prior density distribution:

$$f(\tau_j|\mathbf{y}, \mathbf{x}; \beta) \propto f(\mathbf{y}|\tau_j, \mathbf{x}; \beta)f(\tau_j) \tag{2.16}$$

Given that the prior distribution of parameters is unknown, estimating teacher effects by traditional Bayes methods is unfeasible. However, using the empirical Bayes approach is possible when the model parameters are treated as known, and they are equal to those obtained with the maximum likelihood estimation (MLE). That means $E[\hat{\beta}_j^{MLE}] = \beta_j$, then the posterior distribution can be constructed or estimated.

Thus, the empirical Bayes prediction of teacher random effects will be the expected value of the empirical posterior distribution.

$$\hat{\tau}_j^{EB} = E[\tau_j|\mathbf{y}, \mathbf{x}; \hat{\beta}_j^{MLE}] = \int \tau_j f(\mathbf{y}|\tau_j, \mathbf{x}; \hat{\beta}_j^{MLE})\,\mathrm{d}\tau_j \tag{2.17}$$

Assuming a multivariate normal distribution structure for random effects $\Gamma \sim N(0, \sigma_\tau^2 \mathbf{I}_N)$, and error terms $\mathbf{e} \sim N(0, \sigma_e^2)\mathbf{I}_J$. In addition to the classic assumptions for random effects models such as; strict exogeneity $E[\mathbf{e}|\mathbf{x}, \Gamma] = 0$, and independence of unobserved heterogeneity with respect to other explanatory variables $E[\Gamma|\mathbf{x}] = E[\Gamma] = 0$. We get that the conditional expectation of the vector responses is $E[\mathbf{y}|\mathbf{x}, \Gamma] = \mathbf{x}'\beta + \Gamma$ and its conditional variance is $Var[\mathbf{y}|\mathbf{x}, \Gamma] = \sigma_\tau^2 \mathbf{I}_J + \sigma_e^2 \mathbf{I}_N$.

Under this framework (linear model and joint normality assumptions) we obtain the same estimators of σ_τ^2 and σ_e^2 using MLE. Therefore, the best linear unbiased predictor (BLUP) will be the same as the mean of empirical Bayes posterior.

The empirical Bayes prediction is:

$$\hat{\tau}_j^{EB} = \psi_j(\bar{y}_j - \bar{x}_j \hat{\beta}^{MLE}) \tag{2.18}$$

where $\psi_j = \frac{\sigma_\tau^2}{\sigma_\tau^2 + (\frac{\sigma_u^2}{N_j})}$ is know as the shrinkage factor, which shrinks the MLE estimate towards the mean of prior distribution, in our case 0. The shrinkage factor is also considered as a "reliability" of the $\hat{\tau}_j^{MLE}$ measures obtained from the total residuals of cluster j or $\bar{y}_j - \bar{x}_j \hat{\beta}^{MLE}$.

The empirical Bayes predictions shrinks the maximum likelihood estimation $\hat{\tau}_j^{MLE}$ when the reliability decreases. That means if; (i) the number of observations per teacher decreases, or (ii) the variance between teacher estimates decreases with respect to the total variance of the model. In the first case teachers with less observations will be adjusted toward 0 or prior mean. The second case there is less evidence of heterogeneity between clusters and the estimates adjustment will be made toward 0.

When the reliability increases, the empirical Bayes estimates rely more on the cluster information per group. That happens when the number of observation per teacher increases or the total variance of the model decreases with respect to the variance between group.

However, no matter what type of approach we use to estimate VAMs; whether fixed effects or random effects, it is important to address the main econometric issue to fulfil the exogeneity assumption of other covariates. In the absence of possible or valid instruments, researchers have used one of the two approaches mentioned above.

We next turn to possible causes of failures of the exogeneity assumption of included covariates in equations from **Model 2** to **4**, induced solely by non-random assignments of students to schools/teachers and teachers to schools.

2.3.4 Causes of non-random assignment

One of the main concerns in the above estimations is to do with the non-random assignment of students to schools or teachers, and of teachers to schools. If there is non-random assignment, the unobservable components from equation (2.9) α_i, $\tau_{j,g}$ can be correlated not only with each other, but also correlated with other included covariates. We discuss the effects of non-random assignments on the properties of estimators.

The three possible sources of non-random assignment are:

1. Student-to-school: Even if the allocation of student into schools were made by lottery, there is a potential source of endogeneity because of selection of schools by parents. Parents can also choose to reside in an area where the most suited school is. Authors usually include as many pupil and family background variables as possible to control for this type of selection, assuming there is no other unobserved source of endogeneity driving the student to school allocation.

2. Teacher-to-school: It is more difficult to prove random assignment in this case or finding valid mechanisms to identify non-random allocation of teacher to school. Teachers can respond to principals' offers and change school to get a better matching. The complexity of this matching process can generate sources of endogeneity that can only be avoided if there is a centralised agency randomly allocating teachers to schools. Therefore, authors either assume that assignment is due to time-invariant unobserved factors that can be controlled through including school dummies, or it is simply assumed that there is no correlation between teacher and school unobserved factors.

3. Student-to-teacher (or student-to-classroom): This source of endogeneity has been explored more in depth than the other two. Some authors have proved that several sources of non-random assignment happen in real educational contexts (e.g. Dieterle et al. (2015)). Others have proposed statistical tests to check whether there is evidence of non-random assignment (e.g Rothstein (2010); Kinsler (2012)), where sorting can be identified from the students' or teachers' side.

Students or teachers can be sorted based on observable or unobservable characteristics. However the bias in TE estimators depends on whether teachers are randomly assigned to students or not. For example, even if students are grouped into classrooms but teachers are randomly assigned, estimates of teacher effects might be unbiased if the source of sorting is controlled. On the other hand, if students are randomly assigned but teachers have been sorted based on their ability or effectiveness, then TE estimates might be biased.[12]

Even if assignment depends only on observable and measurable characteristics that can be included in the model, the mechanisms might vary across schools and time. Therefore, there is always a source of potential bias when non-random

[12] One example of non-random assignment on student-teacher matching is; positive/negative ability of students matched to positive/negative effectiveness of teachers.

assignment occurs. The only case where it might be possible to eliminate most sources of endogeneity is when TEs are estimated in an experimental or simulated data framework (e.g.Kane and Staiger (2008); Guarino et al. (2014b, 2015)).

In the following section, we present a representative and specific review of papers related to TE estimations and discuss possible problems with their estimates. Our analysis puts special emphasis on the estimation methodologies and related assumptions.

2.4 Literature review of teacher effects estimators

In this literature review, we consider some of the most recent papers related to teacher effectiveness estimates using **Approaches 2** and **3**.[13]

In the first group (**Approach 2**), authors prefer not to make the crucial assumption of uncorrelatedness and independence of the unobserved TEs with other included covariates such as is proposed by ***A4***. Instead, a set of teacher dummies are included in the model. On the other hand, the second group (**Approach 3**), is based on the assumption that TEs are drawn from a stochastic process and follow a normal distribution with mean 0 and variance σ_τ^2, in addition to ***A4***.

Sometimes, as we will see below, the authors write down a Value-Added model without directly deriving it from a GAF as we did earlier. Hence, it is difficult to know what kind of restrictions they are imposing in their VAM to link it back to our GAF. In particular, our derivation implies a moving-average (MA) error term for **Models 1**, **3** and **4**. However, in the literature reviewed here, most authors do not allow for a (MA) error in their VAM equations.

2.4.1 Authors following Approach 2

Value-Added models for estimating teacher effects became very popular after the Tennessee Value-Added Assessment System (TVAAS) released its measures of teachers' impact on academic performance during the nineties (Sanders and Horn (1994); Sanders and Rivers (1996); Sanders and Horn (1998)). This estimation was based on the test scores obtained by pupils in the Tennessee Comprehensive Assessment Program (TCAP). The objective of the TVAAS was to identify separately the effects of teacher and school on the pupil's learning process, and provide a yearly report of teacher Value Added estimates (publicly available). Teacher VA

[13] We do not include early literature based on **Approach 1** as it has been shown that observed teacher characteristics are not good predictors of TE. Hence, these estimators founded on **Approach 1** have not been used lately.

estimates were generated for teachers who have taught at least one of the specific subjects tested in the TCAP from 3^{rd} and 12^{th} grade.

William L. Sanders (jointly with other authors) carried out a series of studies to test different VAM specification for the TVAAS. Initially, Sanders and Horn (1994) proposed a VAM including historic academic performance, while all student background and school contextual covariates were excluded from the model.[14]

As a response to critics, with respect to the lack of control variables, particularly socio-economic and demographic student background variables, Ballou et al. (2004) extended the original study. These authors assume non-decaying effects of input factors ($\lambda = 1$) as shown in **Model 3**. They also introduce student characteristics such as race and free lunch entitlement at individual and school level. Referring to our VAM specification from equation (2.11), the modified model takes the following form.

$$A_{i,g} - A_{i,g-1} = x'_{i,g}\beta_g + \tau_{j,g} + \varepsilon_{i,g} - \varepsilon_{i,g-1} \tag{2.19}$$

where $\tau_{j,g}$ are considered fixed teacher effects. Inclusion of these fixed effects will take care of possible correlation between $\tau_{j,g}$ and $x_{i,g}$ due to potential non-random assignment of teacher to classrooms. The authors use a two-step estimation strategy to estimate the TE. They first estimate the coefficients β_g by regressing the gains in score $(A_{i,g} - A_{i,g-1})$ on $x_{i,g}$. The residuals per teacher are recovered by averaging over the student scores in her class. Finally these are adjusted by a shrinkage factor. This is the SAR/AR-EB method discussed earlier.

The results of the modified TVAAS model which controls for student background, do not differ considerably from the original TVAAS.[15] Approximately 5,000 teacher Value-Added measures were estimated. Defining a classification of effective and ineffective teachers given a specific criterion, the authors show that the classification made under the two approaches (original and modified TVAS) is highly correlated (above 0.9).

It is important to highlight that the original TVAAS is more flexible than this modified version, as it is able to predict TE with similar results and less data requirement (all included covariates).

Aaronson et al. (2007) estimate TE for Chicago Public Schools. They focus on students and teachers from 9^{th} grade to obtain Value-Added measures. The specification slightly differs from other VAMs where teacher dummy variables $\tau_{j,g}$ are included in the model because 9^{th} grade teachers are mostly specialised in

[14] We will discuss, in detail, some aspects of this paper in the next subsection when we present VAM estimators under **Approach 3**.

[15] The original TVAAS model will be discussed in the following subsection when we review papers which used **Approach 3** to estimate TEs.

subject (SS) teachers. In order to account for some students having two different teachers in two different semesters, they also allow the TE to vary by semester taught in grade g. The indicator variable $d_g = 1$ if there is another teacher (indexed by j') who had taught the student in the same grade. To account for potential correlation between teacher effectiveness and time-invariant school characteristics, the authors also include school FE (s_g).[16]

$$A_{i,g} = \lambda A_{i,g-1} + x'_{i,g}\beta_g + \tau_{j,g} + \tau_{j',g} * d_g + s_g + \varepsilon_{i,g} \tag{2.20}$$

The VAM specification includes a rich set of student, classrooms and neighbourhood covariates which may account for some correlation between TE and observed student characteristics. However, the consistency of the estimator still relies on the strict exogeneity assumption ***A1*** (note the presence of past score in the set of regressors and the omission of $\varepsilon_{i,g-1}$).

The data for the estimation came from a sample of student observations from 1997 to 1999. Using a set of sample selection criteria, such as a minimum number of students per class (15 pupils), the number of maths teachers was reduced from 1,132 to 783. The methodology used to estimate equation (2.20) was OLS and an adjustment of the TEs using a shrinkage factor was applied to the estimated variances of TEs and sample error (σ^2_τ, σ^2_ε).

Regarding the bias generated by potential non-random assignment of teachers to classrooms based on academic performance, Aaronson et al. (2007) compare current test scores dispersion with counterfactual dispersion based on lagged test scores.[17] The comparisons of test score dispersions did not show any significant difference between current real teacher assignment and random counterfactual classes based on lagged test scores. The result suggest that real classes in 9^{th} grade are closer to random counterfactual classes than sorted counterfactual based on test scores.

Authors also compare the ranking of teachers provided by the adjusted (or shrunk) Value-Added estimates from **Approach 2**, with the ranking obtained from the full model estimated by FGLS (**Approach 3**). Both measures were found to be highly correlated (above 0.9), but only when the model included SE

[16] We have suppressed the school sub-index identifier for simplicity.

[17] The authors consider current dispersion of test scores in 9^{th} grade classes and compare them with dispersion of counterfactual classes based on 8^{th} grade test scores. The counterfactual classes refer to classes assigned artificially (or hypothetically constructed) based on 8^{th} grade scores, or based on 6^{th} to 7^{th} and 7^{th} to 8^{th} gains. The perfect sorted counterfactual classes are formed grouping students based on their test score ranking; grouping high performers in one class and low performers in other, in relation to their respective cohort within the school. The random counterfactual classes are randomly constructed among the students within the school-grade cohort, without considering previous test scores or score gains.

either as fixed effects or random effects. That is, the ranking of teachers was very sensitive to whether the school fixed effects s_g were included in the model or not.

There is another stream in the VAM literature which relies on the independence assumptions of TE with respect to observed and unobserved variables and which is supported by $\boldsymbol{A4}$. The next subsection reviews some of the most relevant papers based on this framework.

2.4.2 Authors following Approach 3

The first TE estimations of the TVAAS motivated the discussions and further research on the most recent VAM estimation methodologies. The original TVAAS model proposed by Sanders and Horn (1994) was specified using the following equation (notation adjusted to match ours):

$$A_{i,g} = x'_{i,g}\beta_g + \tau_{j,g} + \varepsilon_{i,g} \tag{2.21}$$

Authors did not have access to observed student characteristics. Therefore, they used district level average scores as a representation of the school system performance and $\tau_{j,g}$ represents the teacher random effect. It is assumed that $\tau_{j,g}$ and $\varepsilon_{i,g}$ are independent and identically distributed (*iid*) $Normal$ random variables.

Under this framework, TE estimates are presented as the best linear unbiased prediction (BLUP), which is calculated adjusting the Maximum Likelihood estimates (MLE) obtained from the model in equation (2.21). This is the MLE-EB discussed earlier.

McCaffrey et al. (2004) present a general model and apply that to a sample of data on 678 student achievement scores in a suburban district in the US. The student scores covered grades 3-5 from five elementary schools. They start with a general model that is similar to our general achievement function (GAF) from equation (2.3). After some modifications, we can rewrite their general specification into our VAM context.

$$A_{i,g} = x'_{i,g}\beta_g + \sum_{r=1}^{g} \dot{\lambda}_r \tau_r + \sum_{r=1}^{g} \ddot{\lambda}_r s_r + \varepsilon_{i,g} \tag{2.22}$$

There are two main differences between equation (2.22) and the GAF in equation (2.3):[18] (i) the omission of individual student level heterogeneity α_i,

[18]Note that there could be more than one teacher per grade r, where $r = 1...g$, but we suppress the teacher sub-index in equation (2.22) to simplify its notation.

and (ii) the inclusion of persistence parameters $\dot{\lambda}$ for teacher effects (τ_r) and $\ddot{\lambda}$ for school effects (s_r), which are different to each other and no restrictions are imposed on how these change over time. This is in contrast with our VAMs, where the persistence parameter λ is assumed to be a common factor parameter which is constant across time. They further assume that the teacher and school level random effects as well as the idiosyncratic error ε, are all *iid* random variables and distributed as $N(0, \sigma_\tau^2)$, $N(0, \sigma_s^2)$ and $N(0, \sigma_\varepsilon^2)$ respectively and are all independent of each other.

The model is then used to test for different restrictions implied by other models that have been used to estimate TEs in the TVAAS program (e.g. Sanders and Horn (1994) and Ballou et al. (2004)). The results obtained from the MLE show that models with and without covariates generate similar teacher value-added estimates obtained from MLE-EB. However, the estimations might be biased if there is non-random assignment of teacher to classroom based on the omitted variables.

The authors conclude that the inference made on TE estimations have to be prudent as they require assumptions that may not hold in real situations. The reasons for the failure may be due to lack of information with respect to non-random assignments of teacher to classroom, and the difficulty of controlling for all possible settings in a non-experimental context.

Similarly, Rockoff (2004) estimates TEs employing data from elementary schools in two districts within a New Jersey county in the US, during the period 1989-90 through 1999-00. The school sample was composed of nearly 10,000 students and 300 general teachers.[19] Because there is only one teacher per class grade-year, TE estimations and classroom impacts cannot be separately identified. The achievement measure variable is the national standardised exam which is taken at district level.

If we modify our VAM from (2.10) in **Model 2**, including the teacher experience function $f(Exp_j)$ proposed by the author, we can rewrite the VAM as

$$A_{i,g} = x'_{i,g}\beta_g + f(Exp_j) + \alpha_i + \tau_j + s_{i,g} + \varepsilon_{i,g} \tag{2.23}$$

The teacher experience function is specified as a monotonically non-decreasing function, which stops increasing after a specific threshold ($\overline{Exp_j}$) of years of experience. Unlike previous VAM specifications, Rockoff (2004) includes fixed effects for school-year pairs $s_{i,g}$ to control for school effects across time, limiting the

[19] General teachers are those who teach all subjects to the same classroom through the grade-year (at least Maths and Language).

teacher Value-Added interpretation to comparisons within schools and within a year. However, in terms of assumptions, the author follows the previous literature and assumes that teacher effects τ_j are random and *iid*. The model is estimated using MLE and the TE predictions are obtained using EB posterior distribution (MLE-EB).

Regarding the source of bias caused by potential non-random assignment of teacher to classrooms, the author relies on official district information which states there is no sorting of student based on achievement or abilities. To support this hypothesis, he also applied a falsification test where classroom dummy variables were not significant predictors of previous pupil's scores. We explore this type of falsification test (Rothstein (2010)) in the next subsection.

Within the RE approach, Jacob and Lefgren (2008) evaluate whether some principals are able to identify and classify teacher effectiveness within schools. The authors estimate a VAM for a school district in a western state in the US. The sample selected from the elementary schools contains 201 teachers who taught core courses (Language and Maths) from 2^{nd} to the 6^{th} grade in the 2002-03 academic year.

A survey was conducted during February 2003, where principals were asked to evaluate teachers in different dimensions such as; motivation, pupil engagement and pupil academic progress. They compare TE estimates with subjective personal evaluations taken from this survey administered by the authors.

Some of the difficulties faced by the authors is that the exam results were not reported as a score, but as a percentage of items answered correctly. In addition, the information collected from the survey of the principals was coded on a subjective scale rate from 1 (inadequate) to 10 (adequate). Both measures, individual attainment and principal's perception, were normalised to have a mean 0 and standard deviation equal 1 in order to make them comparable. The authors use a VAM similar to our **Model 2**.

$$A_{i,g} = x'_{i,g}\beta_g + T'_{j,g}\pi_g + \tau_{j,g} + s_g + \zeta_t + \alpha_i + \varepsilon_{i,g} \tag{2.24}$$

where $x_{i,g}$ corresponds to student characteristics, $T_{j,g}$ are classroom level covariates, $\tau_{j,g}$ represents random teacher effects, s_g school fixed effects, and in addition to our VAM the years fixed effects ζ_t.[20]

Equation (2.24) is estimated by MLE and authors obtain TEs using the EB method (MLE-EB). They found that the correlation between TE estimates from the VAM and the teacher quality perception from the survey to be positive (0.37 Maths, 0.55 Language). In addition, authors propose non-parametric techniques

[20]Note that years t can be expressed in terms of grade g.

to estimate the association between these two different perspectives.

Buddin (2011) follows both RE and FE approach, although he only presents results for the RE approach arguing that results were very similar in both cases. The Value Added measures are estimated with and without assuming independency of teacher effects ***A4***). The author provides Value-Added measures for over 11,000 teachers in the Los Angeles Unified School District (LAUSD). The sample of observations was taken from students and their respective teachers during 2003-04 to 2009-10, from 2^{nd} to 5^{th} grade.

Academic performance is measured using standardised scores at grade level, from the California Standards Tests (CST). The preferred VAM specifications consider the achievement variable in levels and include the lagged achievement scores as a control variable, such as **Model 1** in equation (2.4). No allowance for student specific unobserved heterogeneity was made, and the overall equation error $\varepsilon_{i,g}$ was not represented as a MA error. Two separate equations were specified for estimating TEs and SEs:

$$A_{i,g} = \lambda A_{i,g-1} + x'_{i,g}\beta_g + \tau_{j,g} + \varepsilon_{i,g} \tag{2.25}$$

$$A_{i,g} = \lambda A_{i,g-1} + x'_{i,g}\beta_g + s_g + \varepsilon_{i,g} \tag{2.26}$$

Both models estimate teacher $\tau_{j,g}$ and school effects s_g respectively, using EB method after estimation of the equations by Feasible Generalised Least Square (FGLS). The FGLS procedure involves a 2-stage estimation, where the first stage simply provides residuals from the OLS estimation, and the second stage estimates the structure of teacher (or school) heteroskedasticity, in addition to the error component variance. If assumptions ***A1, A2, A3.2*** and ***A4*** hold, the FGLS estimator would be biased but consistent and a more efficient estimator than a WG estimator.[21]

The author suggests that TE estimates work well identifying broad groups of teacher rather than making one-to-one comparison of teachers. Defining groups at the top and at the bottom of the TE distribution, it is possible to state whether a group of teachers is less or more effective in improving student academic performance.

We reserve a special section within the RE approach to discuss one of the most well-known papers related to teacher effectiveness. Its results have strongly influenced political discussion in the US regarding the importance of providing high quality teachers.

[21] The author also relaxes the teacher effects independence assumptions ***A4*** to estimate Value Added measures. Under this approach, they estimate the VAM by OLS including teacher-year dummy variables. The author found no significant differences in teacher effects estimates applying this method compared to the FGLS.

Comments on Chetty et al. (2014a)

The authors focus on the estimation of TEs with a standard VAM model and check for biases in estimated TEs due to potential non-random assignment of students to classes and also teachers to schools. The data used relate Maths and English test scores in grades 3-8 from a large urban school district, with individual student and teacher identifiers which make it possible to use both as panel elements.

They modify the standard VAM model that considers a fixed TEs to allow for changing TEs over time due to the fact that their data spans a very long time period (1989-2009). Since the model and the estimation methodology is different to the ones we have discussed earlier, we present this model and method in detail now.

We use i and t as the student and time indexes. The teacher index j for student i can be thought of as $j = j(g(i,t))$ where g is grade or year-class used in previous notations. Since the estimation sample uses many cohorts of students, all having many test scores related to different grades over time, we now follow Chetty et al. (2014a) and use t as a general time index to keep the notation simple. Write the VAM equation as

$$A^*_{i,t} = x'_{i,t}\beta + \tau_{j,t} + \epsilon_{i,t} \tag{2.27}$$

$A^*_{i,t}$ is the achievement of student i in year t. The following sets of variables are included in $x_{i,t}$.

Student level variables: (i) separate cubic polynomials in last year's score in Math and English that is allowed to vary by grade (i.e. A student's achievement in grade 4 will be a function of grade 3 Maths and English scores specified as a cubic polynomial and when the student moves to grade 5, there will be a different cubic polynomial that will be a function of grade 4 Maths and English scores); (ii) ethnicity, gender, age, lagged suspensions and absences, indicators for special needs, limited English proficiency, and grade repetition.

Class and school level variables: (i) cubic polynomials in class and school-grade means of last year's test scores in Math and English each interacted with grade; (ii) classroom and school-year means of all other individual characteristics; (iii) classroom size and class-type indicators; (iv) grade and year dummies. The teacher effects are $\tau_{j,t}$ where j correspond to teacher in year t. Then, the TE is specified as

$$\tau_{j,t} = \gamma_j + \mu_{j,t} \tag{2.28}$$

where the above can be thought of as composed of a time-invariant and a time-varying component of $\tau_{j,t}$, which varies from year to year from the long-run mean effect γ_j for each teacher j. They assume that $\mu_{j,t}$ is stationary and define the TE (i.e. the variable of interest), as the $\mu_{j,t}$, allowing the teacher quality fluctuates stochastically over time.

We can now write equation (2.27) as

$$A^*_{i,t} = x'_{i,t}\beta + \gamma_j + \mu_{j,t} + \epsilon_{i,t} \tag{2.29}$$

where $\epsilon_{i,t}$ is an idiosyncratic error which can be thought of as a combination of class room shocks and other shocks and is also assumed to be stationary. Authors also impose the following set of **stationary assumptions**.[22]

$$E(\mu_{j,t}|t) = E(\epsilon_{i,t}|t) = 0, \qquad Cov(\mu_{j,t}, \mu_{j,t+s}) = \sigma_{\mu s}, \qquad Cov(\epsilon_{i,t}, \epsilon_{i,t+s}) = \sigma_{\epsilon s} \qquad \forall t, s$$

Note the absence of $\epsilon_{i,t-1}$ from equation (2.27). The estimation of TE ($\mu_{j,t}$) proceeds as follows:

Step 1: Regress $A^*_{i,t}$ on $x_{i,t}$ and a set of teacher FEs (γ_j) and obtain the average of residuals $\bar{A}_{j,t}$ for the class teacher j in year t.

$$\bar{A}_{j,t} = \widehat{\mu_{j,t} + \epsilon}_{j,t}$$

where $\epsilon_{j,t}$ is the $\epsilon_{i,t}$ averaged over the students in teacher j's class. This is an OLS estimation. This procedure will generate a panel of observations for each teacher over time using the *adjusted* test scores (adjusted for covariates and teacher FE).

Note, the composite error term in this equation is $\mu_{j,t} + \epsilon_{i,t}$. Hence this procedure requires the assumption of strict exogeneity to obtain consistent parameter estimator from OLS. Thus the strict exogeneity assumption needed is

$$E(\mu_{j,t} + \epsilon_{i,t}|x_{i,1}...x_{i,t}...x_{i,T}, \gamma_j) = 0$$

However, this assumption may not hold because although $\mu_{j,t}$ and $\epsilon_{i,t}$ are assumed to be stationary, they are correlated over time. Given the inclusion of prior test scores as covariates, there may be correlation between these variables and these two error terms if there is non-random assignment of students to teachers and/or teachers to schools (i.e. the covariates will not be strictly exogenous).

Step 2: The averaged residuals in year t ($\bar{A}_{j,t}$) is regressed on all the previous years' averaged residuals ($\bar{A}_{j,t-1,...}\bar{A}_{j,1}$) using OLS. The regression prediction is then used as an estimate of teacher value-added (i.e. as an estimate of TE).

[22] Chetty et al. (2014a) include lagged scores as regressors and estimate the model using a panel of individual observations but do not assume a moving average process (e.g. MA(1)) for $\epsilon_{i,t}$ in equation (2.27). The authors assume there is a stationary error process but they do not specify what type of error process is nor how to address this when estimating equation (2.27)

$$\hat{\mu}_{j,t} = \sum_{s=1}^{t-1} \hat{\psi}_s \bar{A}_{j,s}$$

where $\hat{\psi}_s$ are the OLS estimates from Step 2. These will be functions of the cross covariances of all the $\bar{A}$s.

An example of a special case would be when only one previous year is taken into consideration in the estimation of $\mu_{j,t}$. Here, one would regress $\bar{A}_{j,t}$ on $\bar{A}_{j,t-1}$, and $\hat{\psi}_1$ be equal to $\sigma_{A,1}/\sigma_A^2$ where $\sigma_{A,1}$ is the covariance between current and previous periods $\bar{A}$ and σ_A^2 is the variance of $\bar{A}$. The assumption of stationarity presented earlier is crucial here to minimise the number of parameters to be estimated.

They check for biases in their estimator of TE from two sources: (i) due to omission of relevant observable variables such as demographic parental covariates; (ii) due to potential correlation of teacher assignment and omitted student unobservables. To check for the first source of bias, they compare changes in TE predictions between the VAM specification without relevant observed demographic variables and the fully specified VAM.[23] Regarding the second source of bias, the comparison is based on a quasi-experimental framework exploiting teacher turnover to examine changes in individual academic achievement. However, this also relies on the assumption of random assignment of teacher to schools.[24]

It is important to highlight that these validations of whether TE are biased or not are still dependent on the two crucial assumptions: (1) *random teacher-to-student assignment* (same as our ***A2***), and (2) *random teacher-to-school assignment.*

Neither of the two validations tests shows significant levels of bias of TE estimates when both random assignment assumptions mentioned above are assumed to hold. Nevertheless, the challenge is to then prove that the random assignment assumptions actually hold for every educational context where TEs are estimated. Rothstein (2014) has criticised the exogeneity assumption of teacher turnover, showing that Value Added measures are highly sensitive to it.

Apart from the contribution to the technical discussion about the validity of VAMs to estimate TEs, this paper has became so influential academically

[23]The authors have access to household tax records, which allow them to recover information on parental income and mother's age at birth. The results show that after controlling for lagged scores for one period, the bias produced from omitting household income, mothers age birth, or second period lagged scores is insignificant.

[24]When comparing changes in grades in a school where teacher moves, it is expected that differences in predicted TE will be accompanied by proportional gains (or loses) in academic achievement. That is why it is required that teacher turnover is completely exogenous, then changes in mean residual scores per classroom are attributable to differences in Value-Added measures. The results show that this correlation is around 0.95.

and politically because of its striking findings.[25] The authors found substantial differences between being exposed to a highly effective teacher versus an ineffective one. Students assigned to more effective teachers in early primary grades are more likely to attend college, earn high salaries, live in higher socio-economic status neighbourhoods and have higher saving rates. These results have caught the attention of the media and policy makers, supporting the importance of teacher quality issues in the policy debate.

In the following subsection, we present some of the related literature assessing the causality of TE estimates using VAMs, and we show how authors have used some innovative strategies to estimate the bias of TE predictions (e.g. creating an experimental setting, or designing novel simulations of data).

2.4.3 Testing causality of teacher effects

We have discussed how some of the most well-known papers estimate Value Added measures in the teacher effectiveness literature, and how they deal with the required assumptions for each model and estimation approach. These assumptions are difficult to hold in a regular educational context, as it is expected some kind of sorting between students-teachers-schools in most of the educational systems. Therefore, the literature has been recently focused on analysing whether it is possible to attribute causality of TE on pupil academic achievement.

Part of this literature has been developing mechanisms to validate the accuracy of VAMs to predict TEs. Different specifications and estimation strategies have been evaluated under experimental and non-experimental frameworks. Complementary studies have based their analysis on simulated data, finding similar conclusions in terms of estimation validity. These works have arrived at a general consensus that typical VAM specifications which control for enough student, teacher and school characteristics besides previous academic performance provide relatively good predictions of teacher Value Added.

Kane and Staiger (2008) contribute to this analysis as they take advantage of an experimental setting where randomness of student to teacher assignment was essentially assured. Authors suggest that, as $\boldsymbol{A2}$ is hard to hold in a non-experimental framework, it is convenient to check how biased teacher effects estimates would be when using a period of random assignment of teacher to classroom.

To evaluate the level of bias in VAM estimates, the authors designed an experimental framework where teachers are randomly assigned to students within

[25] This paper corresponds to a series of two papers, where the firstly Chetty et al. (2014a) refers to the TEs estimation, and the second, Chetty et al. (2014b), address the issue of long term effects of these primary teachers on labour market outcomes.

schools. The idea was to estimate the Value-Added measures in a regular context and compare their predictions in the experimental framework. It is supposed that any difference in student performance within school-grade would be driven by differences in TEs. If differences in classroom achievement within school-grade are explained by differences in Value Added estimates, the TE estimations would be unbiased.

Taking advantage of a list of teachers applying to a certification programme offered by the National Board for Professional Teaching Standards (NBPTS) in the United States, the authors contacted some of the related principals from a specific county in Los Angeles. They offer the experiment which consisted in randomising the allocation of students to teachers within the same school-grade. Principals were asked to provide lists of students-teacher that in theory were randomly allocated, and they were warned of the chances that researches re-allocate students to teachers (randomly). The reshuffle made by researchers actually occurred for a group of classrooms which were randomly chosen.[26]

The experiment was carried out during 2003-04 and 2004-05, for a selected sample of schools with two classes per grade, and with classes taught by general teachers from 2^{nd} to 5^{th} grade. The selected teachers had to have at least three years of experience and have enough information before the experiment in order to get Value-Added estimates. After applying the selection criteria, the initial sample consisted of 97 pairs of teachers-grade, but only 78 pairs became part of the experiment.

The general VAM specification considers the test scores in levels and gains as alternative dependent variables, and a set of student, classroom, and school characteristics regressors (with and without including unobserved student fixed effects α_i). The composite error term is compounded by teacher random effects $\tau_{j,g}$, classroom random effects c_g, and the idiosyncratic error term $\varepsilon_{i,g}$. Authors proposed four VAM specifications, and the preferred model takes the form of equation (2.30).[27]

$$A_{i,g} = x'_{i,g}\beta_g + \nu_{i,g} \tag{2.30}$$

where $\nu_{i,g} = \tau_{j,g} + c_g + \varepsilon_{i,g}$

Note that the student fixed effects α_i is not included in the model, because it is considered only in a particular specification where all student-level variables are removed from the model.[28] Teacher Value Added is obtained estimating the

[26]All intended to treat schools belong to the Los Angeles Unified School District (LAUSD).

[27]The specification for the dependent variable in gains, and the inclusion of the student fixed effects α_i can be also derived from here.

[28]Author ruled out this specification and they do not even show the Value Added measures obtained with this model.

equation above by OLS and taking the average residuals per classroom (AR estimator) as the TEs estimates. The EB estimates of TE are later constructed adjusting for the individual sampling error at teacher level (SAR/AR-EB). Under this approach $\boldsymbol{A2}$ is required.

To construct the EB of TE, it is necessary to estimate the variance of teacher σ^2_τ, classroom σ^2_c and the error term σ^2_ε. With these estimated variances it is possible to form the shrinkage factor ψ_j which adjusts the OLS averaged classrooms residuals taught by teacher j in period t. [29]

To test whether estimations of TE are unbiased under a non-experimental framework, the authors regress the differences in classroom performance within the 78 selected teachers pairs p which in theory are randomly assigned, on the differences in the Value-Added predictions for teachers j, j' within school-grade.

$$\overline{A}_{j,p} - \overline{A}_{j',p} = \Omega(\hat{\tau}_{j,p} - \hat{\tau}_{j',p}) + \widetilde{\nu}_p \tag{2.31}$$

where $p = 1, ..., 78$. The coefficient Ω represents the responsiveness of variation in academic performance due to differences in TE, and it is obtained from the OLS regression of equation (2.31) with robust standard errors. If the null hypothesis that Ω is statistically different from 1 cannot be rejected, the TE estimator would be unbiased.

The results suggest that for those specifications which control for student, classroom and school characteristics besides lagged scores, the null hypothesis of Ω being different from 1 cannot be rejected. It is important to highlight that the coefficient gets even closer to 1 when the VAM specification includes school fixed effects.

However, most of the Value-Added measures are estimated, and potentially implemented in a non-experimental context. Therefore, it is useful to have an idea of the degree of bias produced by non-random assignment, particularly when non-random assignment of students-to-teacher occurs.

Rothstein (2009) gauges the bias of TE estimator using data from the North Carolina Education Research Data Centre. The analysis is made under two different strategies, which vary depending on the assumptions made to the teacher to classroom assignment process.

In the first case, the assignment process is assumed to be based on observed variables, thus the differences in the magnitude of bias with respect to a baseline model are driven by differences in the VAM specifications. While in the second case, the allocation of teacher into classrooms is assumed to be made by principals

[29] To obtain the empirical Bayes estimates of TE is necessary to weight the residuals average by the shrinkage factor ψ_j which is the ratio between TE variance and total variance $\psi_j = \left(\frac{\sigma^2_\tau}{\sigma^2_\tau + \sigma^2_c + \sigma^2_\varepsilon}\right)$.

based on observed and unobserved data to researchers.

To determine the level of bias when pupil sorting is based on observable characteristics, the author sets the baseline model which explains score gains in grade 5^{th}, with dummy variables for teacher $\tau_{j,5}$, school fixed effects s_5, and controlling for all observed variables of 4^{th} grade and all history of Maths and Language scores. In our VAM context the baseline model will look as equation (2.32).

$$A_{i,5} = \tau_{j,5} + s_5 + x'_{i,4}\beta_4 + \sum_{r=1}^{4} A_{i,r} + \varepsilon_{i,5} \tag{2.32}$$

The author states that if there is random assignment of teacher to classroom conditional on all set of observable covariates of 4^{th} grade, the OLS estimator will report the true unbiased TE estimates. However, due to data constraints it is likely that not all information from the baseline model is available to researchers. Then, simpler VAM specifications are proposed to compare TE estimates with respect to unbiased estimates from the baseline model equation (2.32). Defining $\Delta A_{i,5} = A_{i,5} - A_{i,4}$, we have

$$A_{i,5,} = \tau_{j,5} + s_5 + \varepsilon_{i,5} \tag{2.33}$$

$$\Delta A_{i,5} = \tau_{j,5} + s_5 + \varepsilon_{i,5} \tag{2.34}$$

$$\Delta A_{i,5} = \tau_{j,5} + s_5 + \lambda_4 A_{i,4} + \varepsilon_{i,5} \tag{2.35}$$

$$\Delta A_{i,5} = \tau_{j,5} + s_5 + \lambda_4 A_{i,4} + \lambda_3 A_{i,3} + \lambda_2 A_{i,2} + \varepsilon_{i,5} \tag{2.36}$$

The comparison between standard deviations of teacher effects obtained in the baseline model equation (2.32) and the alternative VAMs, from equation (2.33) to (2.36), is used to assess the estimated level of bias in every case. They found that the first two VAMs, (equations (2.33), (2.34)), explain at most 50% of total bias TE estimates, while the variance in the bias is reduced up to 13% and 3% with respect to estimaties obtained from equation models in (2.35) and (2.36) respectively.

To measure the magnitude of bias when student sorting is based on unobserved characteristics, the author argues that it is necessary to define the set of information available to principals. Then, it requires some calibration across two dimensions: (i) how principals predict student performance, and (ii) how much of this prediction is used to allocate students to teachers.

The VAM specifications from equation (2.33) to (2.36) were estimated under eight different simulated scenarios (from less to more available information for principals). The results suggest that as the amount of information used by principals increases, the variance of the bias also increases for all proposed VAM.

However, the bias is considerably higher for the simplest VAMs (equations (2.33), (2.34)), while for models which control for lagged scores (equations (2.35), (2.36)) the bias is still low compared to the unbiased baseline model from equation (2.32).

In order to complement the VAM validation based on an ideal baseline model, Rothstein (2010) proposes to implement a falsification test which allows him to verify whether the strict exogeneity assumptions ***A1*** hold for the most used VAMs. Thus, to claim teacher causal effects on pupils academic performance, the assumptions ***A1*** and ***A2*** must be satisfied. If there is evidence of non-random assignment of teacher-to-students, it implies that TE would be correlated to some unobserved assignment process, and therefore the exogeneity assumption ***A1*** would no longer hold. Hence, to test whether this is the case for the North Carolina education system, the author estimates typical VAMs including teacher dummy variables of one year ahead $\tau_{j',g+1}$. Then, if there is causal effects of future teachers on current test scores, it would be enough evidence of non-random assignment of teacher to classrooms within school, and then the exogeneity assumption might not be satisfied either.

The three alternative VAMs tested by the author are derived from a GAF, similar to our equation in (2.3) when we assume that the common rate of persistence and geometric lag distribution ($0 < \lambda \leqslant 1$) is constant across grades. In the first and third VAM (equations (2.37), (2.39)) proposed by the author, the persistence parameter λ is actually assumed to be equal 1.

$$\Delta A_{i,g} = \tau_{j,g} + s_g + \tau_{j',g+1} + \varepsilon_{i,g} - \varepsilon_{i,g-1} \tag{2.37}$$

$$A_{i,g} = \tau_{j,g} + s_g + \tau_{j',g+1} + \lambda A_{i,g-1} + \varepsilon_{i,g} \tag{2.38}$$

$$\Delta A_{i,g} = \tau_{j,g} + s_g + \tau_{j',g+1} + \alpha_i + \varepsilon_{i,g} - \varepsilon_{i,g-1} \tag{2.39}$$

As school dummy variables s_g are included in all VAM specifications, the non-random assignment can only be tested at teacher-classroom level, and all TE variations are measured at school-grade level. Thus, it is necessary to consider schools with more than one general teacher per grade.

Taking a sample of students from 3^{rd} to 5^{th} grade, during the period 1998-99 through 2000-01, and using the 5^{th} grade cohort as a reference, the VAM estimation shows that future TE coefficients $\tau_{j',g+1}$ are statistically not significantly different from zero in the first VAM (equation (2.37)). In the second VAM specification (equation (2.38)) the results are similar, although the impact of future TEs in current Maths scores is no longer significant. However, in both specifications serial correlation of error terms which was found to be present will violate the strict exogeneity assumption ***A1***.

The third VAM specification (equation (2.39)) includes controls for individual fixed effects α_i. It is expected that if the assignment of students to classroom is based on fixed ability, then controlling for α_i would potentially satisfy ***A1***. However, the correlation of future TE ($\tau_{j',5}$) on current scores gains (using 3^{rd} and 4^{th} grade) is not constant. These results suggest that the non-random assignment is not fully based on fixed individual ability, therefore ***A1*** is also violated in this case.

Using a fourth, richer VAM as a baseline model that controls for all available previous scores and teacher assignments, the author found that including lagged scores could reduce almost half of the bias when we assume that non-random assignment is based on observable variables only. To measure bias when students are sorted on the basis of unobservables, Rothstein (2010) simulates data for six different possible types of scenarios, from sorting based on purely observed to fully unobserved variables. This analysis shows how close the VAMs ranking of TE estimates is when compared to the true simulated values.

The results suggest that for those scenarios where student sorting was made on historic test scores (up to three grades), but with only one observed, the VAM provide rankings which are correlated between 0.92 and 0.96 with the simulated true ranking. While, the correlation for simplest VAM (without historic data) is only 0.72, and decreases to 0.56 when the sorting is assumed to be based only on unobserved variables.

After these two papers, Rothstein's recommendation is to be cautious when making inferences on TE and their causality implications. Even if it is possible to roughly estimate the level of bias depending on the VAM specification, it is still uncertain what type of teacher to classroom assignment corresponds in every case. Therefore, before imposing strict assumptions on the VAMs, it is necessary to investigate further what type of assignments contained in the data.

In response to Rothstein's falsification test, where none of the VAMs satisfies the strict exogeneity assumption, Kinsler (2012) proposes an alternative falsification test which performs better in small samples, arguing that this is the main reason why Rothstein over-rejects the null hypothesis that future TEs have zero impact on current outcomes.

Kinsler's validation test assumes homoskedasticity of error terms. Under this setting, the estimation test compares two VAM specifications: one restricted *VAM-R* which does not include future TEs, and the other unrestricted *VAM-U* considering them for one period ahead $\tau_{j,g+1}$.[30]

[30]Note that teacher j may vary along years or grades from 1 to g.

$$VAM\text{-}R: \quad A_{i,g} = \eta_g \alpha_i + \sum_{r=1}^{g} \tau_{j,r} + \varepsilon_{i,g} \tag{2.40}$$

$$VAM\text{-}U: \quad A_{i,g} = \eta_g \alpha_i + \sum_{r=1}^{g+1} \tau_{j,r} + \varepsilon_{i,g} \tag{2.41}$$

The parameter η_g represents the grade specific effects that are common across schools, therefore the estimation school-by-school is not feasible any more. Instead, Kinsler estimates both models *VAM-R* and *VAM-U* iteratively, providing initial values for η_g^0, α_i^0 and τ_j^0, minimising the sum of squared residuals.

To determine whether the strict exogeneity ***A1*** holds, the author uses an F test to compare the R^2 from both models. If future TE $\tau_{j,g+1}$ do not have an impact on current scores, the null hypothesis of non difference between the two models cannot be rejected.

To compare between Rothstein (2010) and Kinsler (2012) alternative validation tests, Kinsler implements a Monte Carlo simulation process to generate data for different scenarios. Under this process, it is possible to modify the number of observations per teacher (10, 20, 50, 100) and the type of student sorting (by ability, and by ability and lagged scores). When the number of students per teacher is 20 and the sorting process is designed to be based only on individual ability, the probability of rejection under the Rothstein (2010) methodology is 0.49 while using the F statistic proposed by Kinsler is only 0.04. The accuracy of validation tests improves when the sample increases up to 50, where the rejection probability is 0.11 and 0.06 respectively. These results confirm that Rothstein's test performs well in large sample while Kinsler's test does not have this constraint.

Whether to support Rothstein or Kinsler validation test, it depends on the assumptions we are willing to impose. These results just confirm how sensitive the conclusions can be when testing VAM validation under different specifications, estimators and educational contexts.

However, additional analysis has been proposed in the literature in order elucidate the usefulness and practical contribution of VAM estimating TEs. The papers presented in the next subsection are more focused on understanding the capability of VAM estimators to estimate true TEs and the importance of the educational context identification, rather than robustness based validation tests.

2.4.4 Performance of Teacher Effects estimators

Independent of how reliable the assumptions imposed on the VAM are, it might be possible to measure the accuracy of TE predictions under different type of educational contexts and estimation strategies. Even if VAM estimators are slightly biased, it might be useful, at least, to analyse whether their predictions of TE correctly rank teachers.

In absence of the true teacher effects, and therefore teacher rankings, some authors have tried either to identify non-random assignment settings or to use complex simulated data to assess VAM estimators performance. Building their own data allows authors to define different educational contexts and set a benchmark of TEs consistently estimated. From this benchmark it is possible to evaluate how VAM estimators differ in their TE predictions under different simulated contexts.

Dieterle et al. (2015) highlight the importance of the type of non-random assignments observed in the data when VAMs are estimated. Focusing mainly on student-to-teacher assignment, the authors distinguish between student sorting and the teacher assignment process as two different sources of endogeneity. Students might be randomly grouped into classrooms or discretionally sorted based on observables and unobservable characteristics, such as previous scores and individual ability respectively. On the other hand, teachers might be randomly assigned to classrooms, or sorted based on observable characteristics and unobserved teacher effects.

The authors state that common VAM specifications can easily deal with cases of student sorting when it is based on observable characteristics to researchers. However, the problem gets more complicated when student sorting is based on unobservable variables. Here, typical VAM estimators would perform differently depending on the assumptions imposed on the model specification.

Using data from an anonymous large state in the US, the authors are able to identify different types of non-random assignment and test the capability of some VAM estimators to predict true TEs. They found evidence of non-random assignment within schools estimating the probability that a student is assigned to a particular teacher given a set of 5 observed characteristics.[31] The results show that there is enough evidence of non-random assignment of students to classrooms, and it is worth to consider it in the VAM estimation.

Additionally, Dieterle et al. (2015) also analysed the potential non-random assignment of teachers into specific types of classrooms within schools. Regressing

[31] This approach is taken from Clotfelter et al. (2006) where they also test the independence of assignment based on 6 observable characteristics. An alternative methodology to test for non-random assignment is used by Aaronson et al. (2007), and it consists in comparing scores distribution between current classes and sorted counterfactuals based on previous scores or gains.

some teacher characteristics on classroom average characteristics, the results show some evidence of race matching between teacher-students, and classrooms with higher academic performance in previous years with more experienced teachers. However, most of the student sorting and teacher assignment might vary across schools, grades and years.

The authors separate the data into different scenarios and compare TE estimates for two type of VAM estimators. The methodologies to estimate TE are OLS and MLE estimations, obtaining posterior standard deviation of TE from EB estimations (MLE-EB). The OLS estimator includes teacher dummies to control for teacher fixed effects $\tau_{j,g}$, while the MLE-EB relies on the independence assumption of teacher random effects $\tau_{j,g}$ with respect to other covariates and other unobserved factors, as is stated in ***A4***.

The VAM specifications vary depending on the assumptions made on the rate of persistence λ. The specification proposed here slightly differs from our **Model 1** and **3**.[32]

$$(\lambda = 1) : A_{i,g} - A_{i,g-1} = x'_{i,g}\beta_g + +\tau_{j,g} + \zeta_t + \alpha_i + \varepsilon_{i,g} - \varepsilon_{i,g-1} \tag{2.42}$$

$$(0 < \lambda \leqslant 1) : \; A_{i,g} = \lambda A_{i,g-1} + x'_{i,g}\beta_g + \tau_{j,g} + \zeta_t + \alpha_i + \upsilon_{i,g} \tag{2.43}$$

The estimations show that there are not significant differences between TE estimates from OLS and MLE-EB estimators when there is not student sorting. The high levels of correlation between TE ranking drops when the comparison are made within estimators (OLS, MLE-EB) but using different VAM specifications (e.g. $\lambda = 1$ and $0 < \lambda \leqslant 1$). However, there are some differences between OLS and MLE-EB estimators classifying high quality teachers in the top quantile, if there is evidence of grouping in the sample. Independent from the type of VAM estimator and VAM specification, when there is non-random assignment of student to teacher the risk of misclassification increases. Therefore, it is not straightforward to recommend a particular VAM specification or VAM estimator when there is presence of non-random assignment.

Continuing the investigation of the most appropriate VAM specifications and estimation methodologies, Guarino et al. (2014b) evaluate the performance of VAM estimators on simulated data. Under this approach, it is possible to replicate artificially different non-random assignment scenarios presented by Dieterle et al. (2015).

The simulated data structure allows one to create cases with different com-

[32]This VAM includes a year effects factor ζ_t, and the individual heterogeneity is not affected by the persistence factor λ.

binations of student grouping (no sorting, sorting based on baseline scores, and sorting based on previous scores), and type of student-to-teacher assignment (random, non-random with respect to previous scores).

Guarino et al. (2014b) test estimation methodologies that have been explained in previous subsections, such us OLS, AR, and FGLS, besides others. The VAM specification is shown in equation (2.44) with $0 < \lambda \leqslant 1$.

$$A_{i,g,} = \lambda A_{i,g-1} + T_{j,g} + \alpha_i + \varepsilon_{i,g} - \lambda \varepsilon_{i,g-1} \tag{2.44}$$

.

The authors claim that due to the simulation design of the data generation process, where the number of students per class remains constant, the dynamic OLS estimator (OLS estimator with $0 < \lambda < 1$) provides TE estimates which are proportional to those obtained from the MLE-EB estimator. Therefore, in their paper they only show results for the dynamic OLS.

The analysis compares true TE (from the simulated data) with predicted TE under different VAM specifications and estimation strategies. Based on the assumption that there is a random assignment of students and teachers to schools ***A2***, the authors find that there is not any estimation methodology which is preferable in all contingent scenarios, regarding student sorting and teacher assignment process within schools.

However, in terms of correlation between predicted and real TE, and misclassification of Value-Added estimates above or below the average, the FGLS, AR and OLS (with lagged scores) estimators perform similarly well when the scenarios involve teacher random assignment to classrooms. Even when students are sorted based on previous scores (statically or dynamically) or unobserved heterogeneities, this group of estimators performs well.[33]

Generally speaking, the OLS (with lagged scores), and proportionally the MLE-EB estimators are useful methodologies to predict TE across several scenarios. Despite the lack of consensus regarding the reliability and consistency of most of the VAM observed in the literature, it has been shown with simulated data that there are some VAM estimators which do well when predicting TE under certain conditions, while there are others which predict poorly in most of the settings.

Hence, Guarino et al. (2014b)'s results suggest that it might not be necessary to hold all underlying assumptions of VAM estimators to get good predictions of TE. Therefore, it is not fully recommendable to rely on validation tests for choosing the most appropriate VAM estimation strategy.

[33]It is called static sorting when student grouping is based on a fixed value such a baseline test scores, while dynamic sorting is when grouping keeps changing depending on previous test scores.

In this line, Guarino et al. (2015) extend the analysis initiated by Rothstein (2010) and Kinsler (2012) with respect to validation tests of VAM estimations. Taking advantage of the simulated data employed in Guarino et al. (2014b), the authors implement two typical validation tests on several VAM estimators for all possible student to teacher assignment scenarios.

As true TE are generated from a simulation process, it is possible to check whether VAM estimations with high predictive power are ruled out by common proposed statistical tests. Both tests are based on whether the underlying assumptions of VAM estimators hold. Then, if the strict exogeneity ***A1*** is violated, the VAM would be no longer consistent, and therefore not recommendable to predict TE.

The first validation test presented by Guarino et al. (2015) is an adaptation of the Hausman test, which basically tests for correlation between unobserved student heterogeneity and teacher assignment. The adaptation comes because the testing is based on a Wald test which is robust for serial correlation and heteroskedasticity of error terms. The second validation test is known in the literature as the Rothstein's falsification test (Rothstein (2010)), where the impact of future TE on current academic performance is taken as evidence of non-random assignment of teacher to classrooms.

If the Hausman validation test rejects the FGLS estimator, this means that estimators from the FE approach would be appropriate as TEs are correlated with other covariates. However, under student non-random grouping but with teacher random assignment, the rejection rate for the FGLS is above 60% while its teacher ranking correlation with true TE is higher than the correlation ranking observed for WG estimates. However, when the rate of persistence is assumed to be partial ($0 < \lambda < 1$), the strict exogeneity ***A1*** would be violated with the FGLS estimator, and that implies a full rejection of the FGLS, even though if it has higher prediction power than the WG estimator.

If we apply the Rothstein's falsification test, the FGLS is also rejected in most of the scenarios, repeating the same mistakes observed from the Hausman test when the FGLS performs even better than other consistent estimators of TEs. If there is sorting of students but teachers are randomly assigned to classroom, the dynamic OLS presents lower rejection rates than FGLS, but it increases significantly when student grouping is based on unobserved heterogeneity, although its rejection rates are never higher than the FGLS in this case.

Guarino et al. (2015) conclude that when falsification tests rejects the null of random assignment implying that the VAM estimator is inconsistent, it still might be the case that the TEs estimates rank the teachers correctly.

There is no unique methodology which precisely estimates TEs for all possible non-experimental scenarios. Hence, it is crucial to understand the context where TE are estimated, as it has been shown that it is possible to define effective VAM estimators which predict satisfactorily teacher quality rankings, even if fundamental assumptions of the structural model are violated. Nevertheless, because there is still a degree of uncertainty and variation among estimators when teachers are sorted into groups, we have to be prudent before using TE measures to execute any policy at individual teacher level.

2.4.5 The Teacher Effect estimator choice

Considering the discussion of causality issues of TE estimates and the accuracy of TE estimators presented above, we decided to follow the MLE-EB to estimate our VAMs.

Our decision was based on Guarino et al. (2014a) who compare in detail the MLE-EB, dynamic OLS and AR estimators. They describe the empirical Bayes (EB) estimates as an alternative approach to minimise the variability of the teacher effects predictions by shrinking them towards their mean sample. The shrinking factor increases when the measurement error is larger, usually when we observe less observations or students per teacher.

As it was explained previously, the EB estimation relies on the independency of TE with respect to other covariates and unobserved factors, and the VAM is estimated by Maximum Likelihood estimator (MLE) adjusting later by the proportion of teacher effects variance with respect to the total variance of the model. Nevertheless, it is common to observe in the literature the EB definition in cases where the estimation of teacher effects is based on OLS regressions but adjusted for a shrinkage factor. In that case, it is recommended to estimate TE by MLE-EB.

Particularly, Guarino et al. (2014a) study the difference between dynamic OLS and MLE-EB estimators. They estimate VAM specification from equation (2.44) on simulating data for the same scenarios proposed in Guarino et al. (2014b, 2015). The results show that under random assignment of teacher to classrooms, even when is evidence of student sorting, both estimators perform well but with a small dominance of MLE-EB estimates. In cases where the TEs are estimated with only one cohort, the prediction power reduces for both estimators, but MLE-EB still keep the advantage.

The MLE-EB estimator improves the precision of teacher effects estimates when there are few observations per teacher. Then, as we estimate TEs from single cohorts, we propose to use the MLE-EB estimator conditional on the random assignment of teacher to classrooms which we test later in Chapter 4.

In practice, we carry on the MLE estimation with simultaneous equations, one for Language and other for Language performance. This is due to the fact that we do not have multiple observation of students across years, we have multiple observations of students across subject only. Hence, we have enough variation to simultaneously estimate teacher and school effects with a single cohort.

2.5 Impact of teacher effects: Standard deviation (SD) of measures

In most of VAMs reviewed in previous sections, researchers present TE estimates in terms of dispersion of TE distributions. Although, the concept seems to be quite abstract, its interpretation is simple and intuitive.

If the variability of the TE distribution increases, this means the impact of TE on pupils' academic achievement would be more variable, while if the variability goes to zero there is no difference between teachers impact on students' academic performance. Thus, TE distributions are centred with mean zero and an estimated or predicted in terms of standard deviation (SD).

The Value-Added measures are associated with a specific achievement distribution, which may vary depending on the educational system, evaluation instruments, and examined subjects, among other factors. Therefore, to make TE estimates even more intuitive and comparable across studies, authors generally organise student achievement by percentiles and gauge the impact on the ranking when an average (or a median) student is exposed to 1 SD higher quality teacher.

Depending on the SD of TEs and the distribution of students achievement, it is possible to predict how many percentiles an average student (from the 50^{th} percentile) would move up in the ranking if she were taught by a teacher with 1 SD higher Value-Added, assuming that no-one else is affected by this change within the achievement distribution.

In Table 2.1 we present a summary with some of the estimated SDs of TEs found in the literature. Columns (1) and (2) report the magnitude of predicted TEs for Language and Maths respectively. Column (3) indicates what type of estimator was used by authors to obtain the predictions shown in this table. In Column (4) we indicate whether the TEs are shrunk towards the mean to reduce the point estimate of TEs using the Bayesian shrinkage factor when the variance error is too high. Thus, adjusted SDs of TEs are also called shrunk SDs of TEs.

Although authors could have obtained different Value Added measures from other VAM specifications and estimation strategies in their papers, what we show in Table 2.1 corresponds to their preferred specification and estimator.

Table 2.1: Standard Deviations of Teacher effects estimates in the literature

	Language / Reading (1)	Maths (2)	VAM Estimator (3)	Type of teacher VA estimate (4)
Aaronson, Barrow, and Sander (2007)	-	0.13	OLS	*Shrunk SDs of TEs*
Chetty, Friedman and Rockoff (2014a)	0.12	0.17	OLS	*Shrunk SDs of TEs*
Buddin (2011)	0.16	0.25	FGLS	*Shrunk SDs of TEs*
Hanushek and Rivkin (2010a)	-	0.11	OLS	*Shrunk SDs of TEs*
Jacob and Lefgren (2008)	0.12	0.26	MLE-EB	*Shrunk SDs of TEs*
Kane, Rockoff and Staiger (2008)	0.13	0.10	OLS	*Shrunk SDs of TEs*
Rockoff (2004)	0.08	0.10	MLE-EB	*Shrunk SDs of TEs*
Rothstein (2009)	0.11	0.15	OLS	*Shrunk SDs of TEs*

Note:(i) Estimated teacher effects are in terms of standard deviations (SD) on student achievement. Most of them present estimations of teacher effects on both Reading/Language and Math standardised scores, columns (1) and (2) respectively. **(ii)** The VAM estimator, from column (3) refers to the estimation methodology employed by these authors to estimate their own Value-Added models (OLS: Ordinary Least Squares; FGLS: Feasible Generalised Least Squares; MLE-EB: Maximum Likelihood with empirical Bayes teacher effects posterior distribution. **(iii)** The teacher Value-Added (VA) estimate is generally adjusted of shrunk to correct for measurement errors of estimated teacher effects. Thus, in column (4) we show whether the estimated SD teacher effects were shrunk or unshrunk.

The Value-Added estimations slightly differ among the studies, with an approximated observed range of SD teacher effects between 0.10 and 0.30. Mostly all authors propose that TE are higher in Maths than Language performance, with only few exceptions (e.g. Kane and Staiger (2008)), where the impact is stronger in Language than Maths, although with a narrower gap of SD TE between subjects.

All reported SDs of TE distributions imply that a 1 SD higher teacher Value Added would increase student test score by the estimated SD of TE in Language or Maths. In Chetty et al. (2014a) for example, an average student is exposed to 1 standard deviation higher quality teacher, it is expected that he or she improves by 0.12 SD and 0.17 SD in normalised Language and Maths test score, respectively.

Despite the differences in SDs of TE predictions and potential level of bias, it is widely accepted that unobserved TEs have a significant impact on student performance. The estimation of its magnitude and the construction of true teacher ranking is still under discussion, as it is not possible to recommend a unique VAM specification and estimation strategy.

2.6 Conclusion

In this chapter we have motivated the implementation of Value Added models in estimating teacher effects in different educational contexts. Starting from a cumulative education production function, we derived a typical general achievement function (GAF) from which it is possible to construct most of the VAM specifications observed in the literature.

We presented the required assumptions depending on what type of VAM specification and estimation strategy is employed to estimate teacher Value Added. We identified three different approaches to estimate TEs from VAMs: **Approach 1**, which relies only on observable teacher characteristics; **Approach 2**, which uses teacher dummy variables to identify teacher effects; and **Approach 3** which assumes TEs are drawn from a random distribution. However, the consistency of TE estimates will depend on whether strict exogenety assumptions, random student-teacher-school assignment and on the type VAM specification and estimator used.

Even if VAMs are correctly specified, it is difficult to test whether VAM assumptions ***A1-A4*** actually hold in non-experimental environments. Most of the educational contexts are exposed to many source of endogeneities due to potential non-random assignment among students, teacher and schools. Therefore, there is an important part of the literature which has been devoted to validate the accuracy of VAMs to predict TEs.

Most of the VAM validation studies conclude that under random assignment of teachers to classrooms, there are no considerable differences between VAM estimators performance, once VAM specifications control for enough student, teacher and school characteristics (including previous pupil attainment). Nevertheless, most of Value Added measures are estimated and potentially implemented in a non-experimental context.

Recent discussion has centred on how VAM estimates perform under non-random assignment. To test VAM accuracy we have checked several strategies used in the literature such as, experimental (Kane and Staiger (2008)) and quasi-experimental frameworks (Chetty et al. (2014a)), statistical validation tests (Rothstein (2010); Kinsler (2012); Guarino et al. (2015)), and evaluations using simulated data (Dieterle et al. (2015); Guarino et al. (2014b, 2015)).

Under an experimental framework, Kane and Staiger (2008)) have shown that even though TEs are estimated with non-random assignment, their predictions are significantly unbiased when they are tested in a random experimental environment. On the other hand, Rothstein (2010) recommends caution with inferences from teacher effect estimates after implementing the falsification test on several VAM specifications, and concluding that, at least in the North Carolina dataset, the strict exogeneity assumption is regularly violated.

Alternatively, Kinsler (2012) proposes a simpler version of Rothstein's validation test, assuming homoskedasticity of error terms. Replicating the sample of Rothstein (2010), Kinsler (2012) confirms that Rothstein's test performs well only for large samples of students per teacher. Whether to choose Rothstein or Kinsler's validation test it depends on the assumptions imposed. However, it seems that statistical validation tests are not enough to rule out the VAM estimations even if exogeneity assumptions do not hold.

Regarding the capability of VAM predicting true TEs, it is useful to check the results obtained from simulation exercises carried on by Dieterle et al. (2015). The authors showed there are no significant differences in VAM estimators performance when random assignment of students to teachers happens, but the construction of teacher rankings might differ when there is evidence of non-random assignment. Thus, the preferable VAM specification and estimation strategy will depend on each educational context.

Also using simulated data, Guarino et al. (2014b) suggest that general dynamic OLS and MLE-EB estimators are useful methodologies to predict teacher effects across several scenarios, and it is not completely advisable to rule out VAM estimations just based on typical validation tests. Similarly, Guarino et al. (2015)) conclude that it is possible to find an useful VAM estimator which satisfactorily predicts teacher quality rankings, even if fundamental assumptions of the structural model are violated.

After extensively reviewing the teacher effectiveness literature, we are able to address our VAM estimation of teacher effects on pupil academic performance for the Chilean school system. To our knowledge, this is the first attempt to estimate teacher effects in Chile. Besides, we will also estimate school effects, analysing their stability and trajectory for a specific period of time.

Chapter 3

Data, cohort samples and performance measures

Abstract

The objective in Chapter 3 is to describe the multiple sources of the Chilean school system data bases from which we construct our own Student Panel Dataset (SPD). Using the SPD, we take cross section cohorts and select panel samples which we match with richly informative and high relevant data, such as the Chilean National Examination (Simce), school marks record, teachers register and school staff administrative data bases.

The second part of this chapter presents the available performance measures: (i) National standardised exams scores, and (ii) school marks by subjects. Despite the constraints in the availability of Simce scores for every grade, we show that school marks are good proxies to replace standardised examination scores when by construction this information is not available.

After reading this chapter, the reader would be able to understand how the available information is organised, and how we construct our sample cohorts with the complementary data necessary for further analyses throughout the thesis.

Contents

3.1 Introduction

The current chapter describes different sources of data which we use in analyses throughout the thesis. This information was provided by the Chilean Ministry of Education (Mineduc) and we discuss here how they relate to each other. The data sets consist of administrative data and the yearly Standardised National Examination (Simce), which besides the individual pupil examination scores, provides survey data responded by parents and teachers of intake students.[1]

First, we show all the available data sets and we explain in detail how we construct the Student Panel Dataset (SPD), from which we select specific cohorts for our analyses. Other relevant data bases such as the Simce dataset, with its complementary bases, are described in depth.

Most studies related to school or student performance use standardised examination scores as a valid and comparable instrument for academic achievement across schools and years. However, in Chile the Simce exam does not provide scores for all students every year as it is only taken in some specific grades.

The main analyses are presented in Chapter 5 and Chapter 6, where we estimate Teacher Effects (TEs) and School Effects (SEs) from Value Added Models (VAMs). These estimation require at least two consecutive annual test scores for each student. Since, the Simce exam is only taken in particular grades (non-consecutive), we use in Chapter 5 the 4^{th} grade 2005 cohort's Simce scores as our basic dependent variable, and the 3^{rd} grade 2004 school marks for the lagged scores required by our VAM specification.

[1]It is important to highlight that many of the data sets here mentioned included data from to the latest delivery of information from the Chilean Ministry of Education (Mineduc) in November 2013.

The analyses require two validations: (i) the school marks in grade 3 are good proxy for the students' attainment (i.e standardised school marks and standardised Simce score are correlated. Refer to Sections 5 and 6 of this chapter); (ii) there is sufficient evidence to ensure that there is random assignment of students to teachers or classes based on previous school marks (i.e comparing school marks distributions between 3^{rd}-4^{th} grades and 4^{rd}-5^{th} grades. For details see Chapter 4).

The correlation analysis between standardised school marks (at school level) and standardised Simce scores is carried out using the following cohorts: 4^{th} grade 2005, 4^{th} grade 2007 and 4^{th} grade 2009. We use the same 4^{th} grade cohorts as reference to verify the random assignment of students to teacher based on schools marks. Although, we also consider their respective school marks for 3^{rd} and 5^{th} grade.

In addition to all student registers and their complementary data, we use the available performance datasets (e.g. Simce scores and school marks) not only to estimate our VAMs, but also to validate the proxies employed, and assumptions imposed to consistently estimate TEs and SEs in Chapters 4 and 5. See Table 3.1 a summary description of selected cohorts and the usage of performance data bases.

Table 3.1: Usage of performance data bases by chapters and cohorts

Chapter	Descrption	Simce scores	School Marks
Chapter 2	*Correlation analysis between school marks and Simce scores*	4th (2005) 4th (2007) 4th (2009)	4th (2005) 4th (2007) 4th (2009)
Chapter 3	*Non-random assignment evidence*	- - -	3rd (2004); 4th (2005); 5th (2006) 3rd (2006); 4th (2007); 5th (2008) 3rd (2008); 4th (2009); 5th (2010)
Chapter 4	*VAM estimations I*	4th (2005)	3rd (2004)
Chapter 5	*VAM estimations II*	4th (2005) 4th (2006) 4th (2007) 4th (2007) 4th (2009)	3rd (2004) 3rd (2005) 3rd (2006) 3rd (2007) 3rd (2008)

Notes: **(i)** In Chapters 2, 3 and 4 we use standardised Simce scores and Standardised schools marks at school levels. **(ii)** Chapter 4 we use school marks in absolute terms.

3.2 Presenting the available data sets

The available data are categorised in two main groups: (1) the administrative data bases, and (2) the National Examination (Simce). All data have been provided by the Study Centre of the Ministry of Education (Centro de Estudios - Mineduc).

3.2.1 The administrative data set

The following data bases are composed of information required of all schools officially registered in the Mineduc at National level. The information is collected on a yearly basis which makes it possible to generate longitudinal data sets at student, teacher and school levels. The sources of data are listed here.

1. Enrolment Data Base
2. Performance Data Base
3. Student Marks Data Base
4. School Directory Data Base
5. Teachers Data Base
6. School Staff Data Base

The Enrolment Data Base (DB), Performance DB and Student Marks DB can be matched using a unique and masked student ID number (Mrun). They can be merged to Schools Directories DB at school level by a unique school code (RBD). Simultaneously, the Teachers DB and the School Staff DB are matched by a unique and also masked teacher ID number, and both can be linked to school and student data sets through the RBD, grade, and class letter variables. Therefore, we identify teachers per subject in every single grade and class per school.

3.2.2 The National Examination (Simce) dataset

National examinations scores by each pupil are considered a key factor for assessing academic achievement, they provide a standardised evaluation which is comparable among schools and across years.

In particular, the Chilean National Examination (Simce) is taken every year but not for every grade. Initially, the grades in which the Simce was taken were rotating among 4^{th}, 8^{th} and 10^{th} grade. That means only one grade per year used to take the Simce. However, since 2005 onwards, the 4^{th} grade started being tested every year, and only 8^{th} and 10^{th} were rotating, one year each.

Previously, Language and Maths were the only subjects to be tested, in recent years new subjects were incorporated such as Science and History.[2] In addition, since 2012, the 2^{nd} grade started being evaluated every year in Language,

[2]We refer as Language to Spanish grammar and reading comprehension, and we keep the same definition for the school marks and teacher's subjects.

and from 2013 the Language and Maths exams will be constantly applied to 4^{th}, 6^{th} 8^{th} and 10^{th} grades. Our available information is from 2003 to 2012.

The Simce dataset is also composed of individual questionnaires (or surveys) which are answered voluntarily by parents and teachers. So, besides the individual scores, the available information coming from parent's questionnaire provides pupil socioeconomic background, such a parental levels of education and household level of income. While the information obtained from teacher's questionnaire relates to teacher characteristics, teacher perceptions and teaching strategies.

In sum, the list of data sets related to the Simce examination is the following;

1. Individual Scores
2. School Directories
3. Municipality Directories
4. Parental Questionnaire
5. Teachers Questionnaire

3.3 Generating the student panel and the data cleaning process

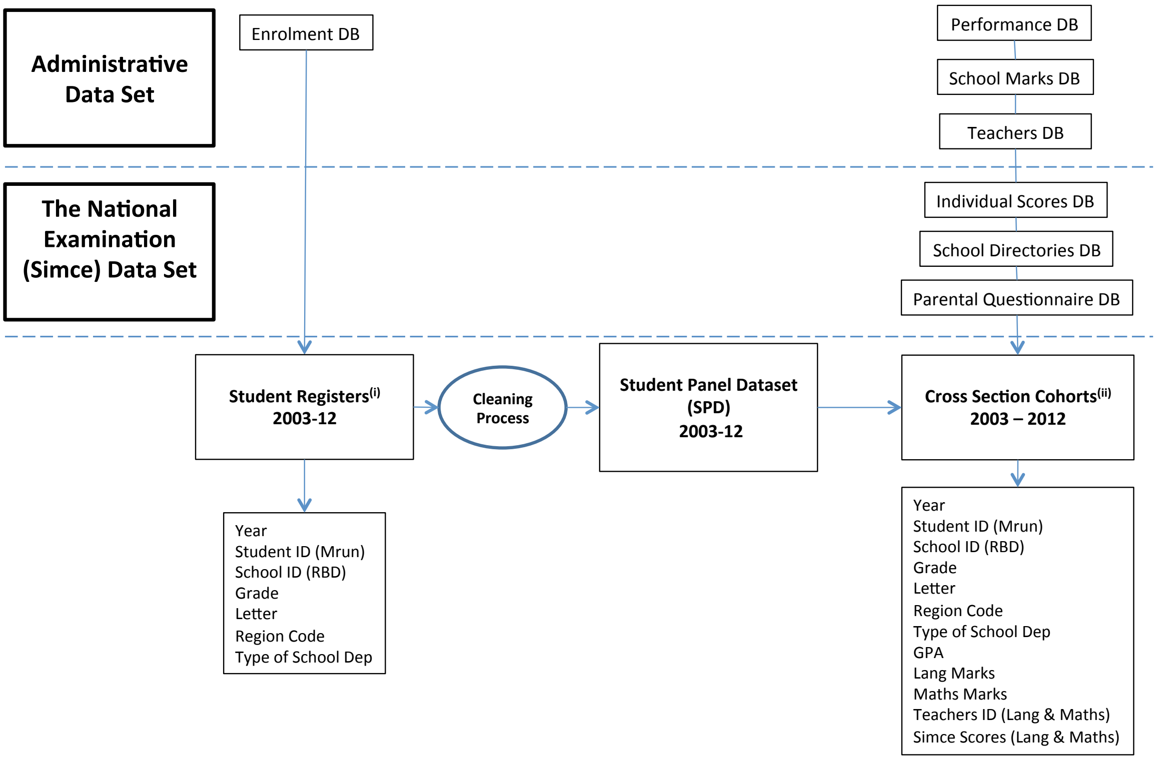
We process the data in a way that most closely matches our research interests. Initially, we construct the Student Panel Dataset (SPD) with the basic student's register, and we add student school marks, teacher identifiers, and the corresponding National examination results.

Given the available administrative data set we assemble the main structure of the SPD based on the Enrolment DB. Although the availability of this data base is from 2004, we use the Performance DB for the year 2003, which slightly varies from the Enrolment DB.

Both data bases (Enrolment DB and Performance DB) are very similar, but they are submitted in different periods of the year and are used by the Mindeduc for different purposes. Although they do not have a perfect matching using the student ID (Mrun), these data bases can be considered as substitutes. Thus, we start the student panel from 2003 taken from Performance DB and we continue appending yearly the Enrolment DB up to 2012.

However, before combining these data bases, which in theory have the same formats, we need to work with them and standardise all the aggregated information for every year. Once the yearly data are merged in a single data set, we are able to start the data cleaning process of our SPD. In Figure 3.1 we show a flowchart representing the dataset generation process from the original sources to the SPD and posterior cross-section cohorts.

Figure 3.1: Processing Data Sets - Flowchart



Notes: (i) To construct the Students Register we use the Enrolment DB from 2004 to 2013, and the Performance DB for 2003 as the Enrolment DB was not available. **(ii)** We use cross section cohorts for the VAM and we also create Mini Panels for particular years and grades (i.e. In Chapter 4). Other data bases (or variables) can be merged on request.

Basically, from two main sources: (i) Administrative dataset and (ii) Simce data set, we put together yearly information to apply the data cleaning process that allows us to construct the final SPD for the period 2003-2012. However, most of our analyses are based on cross-section cohorts, which are selected from this SPD.

3.3.1 Cleaning the Aggregated Student Register (2003 - 2012)

As we identified some cases where students had more than one observation per year, meaning in general more than one school per year, we had to clean the data designing a selection criteria that assign only one school per student-year. This process enables us to construct our student panel data (SPD).

The total number of initial observations was 27,561,043 where a group of 715,334 observations belonged to students who had at least one year of duplicated registers. The group with assignment problems, or more than one school per student-year was classified as "*Bad IDs*" group and it was taken aside to apply the cleaning process.[3] Even if it just accounted for 3% of the total data set it was necessary to define coherent school selection criteria corresponding to possible reasons of pupil's duplicated observations per year (e.g. a school did not update a student register after moving to another school).

In order to generate rational selection criteria, we ruled out all the individuals who had in one year at least three or more registered schools. This group represented less than 0.01% of the initial number of observations. Then, we continued working only with those individuals who had one duplicated school registered in at least one year. The total sum of observations for this group is 583,097. However, if we only focused specifically on years with duplicated cases the observations are reduced to 98,212, which are the final cases where we apply the school assignment criteria.

The logic of the school allocation criteria basically consists in assigning a school ID based on backwards induction. If a student had more than one school in a particular year t, we look forward to the following year $t+1$ to check whether the student is enrolled (without duplicated registers) to one of the schools registered in t. In case there is a match, we leave the school which only appears up to year t, as we can identify the school movement from t to $t+1$. Unfortunately, we do not know the point in time at which the pupil changed school during year t.[4]

When the backwards rationality cannot be applied because there is not a clear sequential school transition from one year to the next, we had to apply a random assignment process.[5] However the number of cases randomly assigned was 25,103 and it represents only 0.01% of the student panel. Finally, we ended up with a panel of 27,428,806 observations, distributed across ten years and twelve grades.

[3]We name "*Bad IDs*" to those students who have at least one duplicated observation per year, in any year over the period observed. Then we take apart all their history, with and without duplicated cases to reconstruct a proper panel with only one observation per student-school-year. Details of the school selection criteria are available in Appendix 3.1.

[4]The backward induction selection criteria is only applied to duplicated with 2 schools per year. When a pupil is registered to 3 or more schools, we dropped the register from the dataset.

[5]The random assignment process consists in choosing one of the two possible schools, setting a random uniform distribution for both and picking either the school below or above the median.

3.3.2 Merging with complementary data sets

Due to capacity restrictions, we are going to work with our constructed Student Panel Dataset (SPD) as the main panel, from which we obtain the cross section cohorts and mini panels used for the analyses in later Chapters. However, additional information will be added depending on the interests and purposes of potential investigations. The idea is to define representative cohorts to work with and merge them with all necessary information.

The complementary data sets can be linked by student ID (Mrun), school ID (RBD), and when we refer to the teacher data bases we also use grades (CodGrade) and classroom identification (Letter). Keeping the student panel structure we just add variables per student-year for those selected cohorts used for analyses and estimations.

Before merging the SPD with the rest of the data base, it is necessary to clean-up and process the data for every period. In the following section, we show the composition of each complementary dataset with a brief description depending on the level at which we have processed the data bases up to now.

3.4 Describing the data sets and student panel

We focus on two groups of data bases and we use them to different extents depending on the research needs. As we mentioned earlier, the two groups are classified as; (1) the administrative data set, (2) the National Examination (Simce) data set.

The SPD is exclusively created from the administrative data set, using mainly the Enrolment and Performance DBs. From the SPD we select the cross section cohorts and mini panels, when it is required. In the following subsections, we describe the two groups of data bases in terms of variables composition. In Appendix 3.2, we provide a summary table of the data contained in each of the data sets with a definition of each variable and their availability of the variables over the years.

3.4.1 Description of the administrative data set

The administrative data set refers to the source of information related to all official registers about schools, teachers and students. The data itself are collected for administrative purposes on a yearly basis. These data bases were not particularly designed to be a longitudinal panel but given their availability over the years we are able to construct our own SPD.

All available information is provided at national level and it is possible to be linked to each other by student ID (Mrun), school ID (RBD) and teacher ID. Within the administrative data set, we classify six data bases described below. Some of variables overlap across data bases, that is why we aggregate all of them in a unique list referring to the administrative data set variables. See Appendix 3.3, a description of the list of variables observed in the administrative data set and their availability through the period 2002-2013.

Enrolment data base

This is a yearly register of all the students enrolled in the Chilean schools system. Every type of school has to provide the list of students officially enrolled to the Ministry of Education by the end of April (the academic year regularly starts the first week of March and finishes in middle December).

Performance data base

The Study Centre of Mineduc manages a yearly data base using the final academic report submitted by all schools at the end of the academic year. Here, it is possible to recover the final academic status, the grade point average (GPA) and school attendance.[6]

Given the information contained in this Performance DB, it is more likely to find duplicated registers per student-year than in the Enrolment DB (due to possible changes in GPA or final status). The evidence also confirms this conjecture, the rate of observations with duplicated schools per student is higher if we use the Performance DB, than the Enrolment DB when we use them as the reference data base to construct the SPD.

School marks data base

This is the official register of the whole aggregated student's school marks.[7] Every student has as many observations as school subjects taught per year. We recover the registers related to Language and Mathematics subjects (the rest of the subjects are available as well). It is also possible to match these records with the teacher's subject area reported in the teachers data set.

[6]The grade point average (GPA) correspond to the average mark of all final subjects' mark by the end of the academic year.

[7]All schools report the final mark for every student enrolled. The final mark correspond to the average mark obtained for the whole academic year in each specific subject.

School directory data base

In parallel to the individual student register, the Mineduc has also constructed a yearly register of all schools in the Chilean education system. Most of the variables contained in this school register data base are already included in the Enrolment DB or Performance DB. What is exclusively in the School directory DB is the school phone number and address.

Teachers data base

The teacher records are organised in yearly files where each observation identifies a teacher ID with the subject and the specific classroom taught. This means, it is very common to find multiple observations per teacher ID. For example, general teachers have different registers for every subject (Language, Maths, Science, etc) taught in the same class. On the other hand, the majority of teachers from 5^{th} grade onwards are subject specialist (SS) teachers. We call SS teachers to those teachers who teach a specific subject to more than one class in the same grade, or in multiple grades and schools.

For matching teacher IDs to the SPD we have to separate the teachers data base in two; Language teachers and Maths teachers. Then, we merge those data bases independently to a specific cohort (using school ID, grade, and class letter). We identify a SS teacher when for the same classroom we observe to different teacher IDs associated, one for Language and another for Maths.

School staff data base

This data set identifies all teachers and managerial staff with a contractual relationship in a school. Every employee has associated a unique and masked ID provided by the Mineduc, which in case of teachers is the same as the one used in the Teachers DB described above.

The observations are organised by job positions, identifying the role of the employee in the school. It is possible to find more than one observation per individual as one teacher could have more than one role in the same or a different school. However, the teacher's ID key variable allow us to enrich the SPD with some observable teacher characteristics obtained from this data base.

Apart from having available individual characteristics such as; gender, date of birth and the experience in the education system, the fact we have access to the role of the individual in the school, allow us to identify principals and some of their characteristics as well.

3.4.2 Description of the National Examination (Simce)

As we have mentioned earlier in this chapter the National Examination (Simce) is a standardised test which started being taken in 4^{th}, 8^{th} and 10^{th} grades in alternate years. Recently, it has been extended to 2^{nd} and 6^{th} grades and with less grade testing alternation. The main tested subjects are Language, Maths, History, Biology and Social Sciences, depending on the grade where pupils are taken the exam. However, in our investigation we are focused only on Maths and Language scores.

In absolute terms the scores have an approximated mean of 250 points with a standard deviation 50 points. Every year and for every subject, the Mineduc defines specific cut-offs to classify the scores into three achievement levels (Low, Intermediate, and Advanced). In our investigation we standardise scores at the year national level.

Despite this information is publically, the individual scores are not accessible to everyone. The Mineduc provides the results at school and municipality level, but only researchers are able to ask for scores at individual level, conditional to develop a research project.

Individual Scores and Directories data bases description

The following tables present the available variables for these two data bases during the period 2003 - 2009. Both data bases provide information to our SPD regarding the individual performance and school level characteristics.

Table 3.2: Variables Description - Individual Scores DB

Individual Scores - From 2003 to 2009 (*Alumnos Mrun)*	
RBD	School ID: key code for schools
Mrun	**Unique identification number per student (Pupils kees the same ID or key number for the whole period)**
GenMrun	Student gender (0 Male; 1 Female)
CodClass	Class code (It contains the grade and classroom letter)
Letter	Identification the class in the same grade-school (e.g. from A to D, depending on the number of students enrolled)
LangScore	Language individual scores
MatScore	Maths individual scores
BioScore	Biology individual scores

Note: We use the Mrun variable (student ID) to match with the rest of the data sets.

Table 3.3: Variables Availability - Directories DB

Directories (School Level) - From 2003 to 2009 *(Establecimientos)*	
CodRegRBD	Code of the region where the school is set. Regional Politic Division of Chile (1-13) and (1-15 From 2007)
CodMunRBD	Code of the municipality where the school is set
NameRegRBD	Name of the region where the school is set
NameMunRBD	Name of the municipality where the school is set
RBD	**School ID: key code for schools**
NameRBD	Name of the school
StudentSimce	# of Students with examination scores (Students who take the test this year at the school)
CodDep	Type of school dependence (0 Municipal; 1 Private Voucher School; 2 Unsubsidised Private School)
RuralRBD	Geographic area where the school is set (0 Urban; 1 Rural)
SocioecGrpRBD	Socioeconomic group of the school (0 Low; 1 Mid-Low; 2 Middle; 3 Mid-High; 4 High)
IVE	Vulnerability index (from 0 to 100)

Note: We use the RBD variable (school ID) to match with the rest of the data sets.

Parental Questionnaire description

Students' parents who took the Simce exam are also invited to respond to an individual questionnaire. The aspects considered in the questionnaire are related to parental education level, household income and satisfaction level with the school.

However, over these years different sets of questions have been asked in parent's questionnaires. We have made a first effort to identify the main variables which were always asked and we made them equivalent for the whole panel period. A brief description of them and their availability is shown in the two tables bellow.[8]

Table 3.4: Variables Description - Parental Questionnaire DB

Parents Questonnaire - From 2003 to 2009 (*Encuesta de Padres)*	
RBD	School ID: key code for schools
CodClass	Class code (It contains the grade and classroom letter)
Mrun	**Unique identification number per student (Pupils kees the same ID or key number for the whole period)**
(Q1) Student Residence	Municipality where the student lives
(Q2) Monthly Household Income	(0 less than £120; 1 from £120 to £240; 2 from £240 to £480; 3 more than £480)
(Q3) Mother Education Level	(0 Prim inc.; 1 Prim comp; 2 Sec inc; 3 Sec comp; 4 Coll inc; 5 Coll comp; 6 Uni inc; 7 Uni comp; 8 Postgrad)
(Q4) Father Education Level	(0 Prim inc.; 1 Prim comp; 2 Sec inc; 3 Sec comp; 4 Coll inc; 5 Coll comp; 6 Uni inc; 7 Uni comp; 8 Postgrad)
(Q5) Monthly Education Expenditure	(0 none; 1 less than £10; 2 from £10 to £50; 3 more than £50)
(Q6) Pre-kinder Attendance	Student attended non compulsory nursery school (0 No; 1 Yes)
(Q7) Selection	Student was affected by selection when applied to the school (0 No; 1 Yes)

Note: We use the Mrun variable (student ID) to match with the rest of the data sets.

Table 3.5: Variables Availability - Parental Questionnaire DB

Availability of Variables in the Data Set by Year/Grade											
Year	**2003**	**2004**	**2005**	**2006**		**2007**		**2008**		**2009**	
Grade	**10th**	**8th**	**4th**	**4th**	**10th**	**4th**	**8th**	**4th**	**10th**	**4th**	**8th**
RBD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mrun	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
(Q1) Student Residence	Yes	Yes	No	No	No	No	No	No	No	No	No
(Q2) Monthly Household Income	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Q3) Mother Education Level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Q4) Father Education Level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Q5) Monthly Education Expenditure	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No
(Q6) Pre-kinder Attendance	Yes	Yes	Yes	No	Yes	no	Yes	Yes	Yes	No	No
(Q7) Selection	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: We use the Mrun variable (student ID) to match with the rest of the data sets.

[8] In Appendix 3.4, we explain in detail the cleaning process of the parental questionnaire data set.

Teacher Questionnaire description

Teachers from classes taking the National Examination are asked to answer questions about their professional career, topics taught and strategies applied.

In 4^{th} grade, only one teacher per classroom responds to the questionnaire as it is expected that all subjects are taught be a general teacher. On the other hand, in 8^{th} and 10^{th} grades the questionnaire is answered by a Language and a Maths teacher separately.

As in the parental questionnaire, the questions are not always the same across years. Thus, it is necessary to clean the data and make the relevant variables equivalent or comparable across years. After going through this process we get to the final set of available and processed variables. To see the description and the availability of these variables, see Appendix 3.4.

However, in this data base we are not able to identify the teacher's ID because the questionnaire is anonymous, but we still can assign it to the respective 4^{th} grade general teacher using the key variables; school ID (RBD), grade (CodGrade) and classroom identifier per grade (Letter).

3.4.3 The Student Panel Dataset (SPD) description

It is important to know the nature and extent of the data set we are working with. We now present a basic description of the SPD formed with the Enrolment and Performance DBs. The SPD will be merged to other data bases depending on the research objectives.

Table 3.6: Number of Students by year and grade

	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade	6th Grade	7th Grade	8th Grade	9th Grade	10th Grade	11th Grade	12th Grade	**Total**
2003	267,338	280,082	278,040	290,172	295,054	299,644	306,242	293,137	194,479	169,349	117,630	92,050	**2,883,217**
2004	262,237	266,394	269,678	278,501	290,053	290,688	295,701	290,226	198,666	172,781	125,861	104,267	**2,845,053**
2005	259,705	259,676	265,856	268,162	284,519	288,700	289,509	284,627	205,278	181,017	130,937	111,094	**2,829,080**
2006	254,492	258,553	259,625	265,681	274,452	284,014	288,007	278,306	207,301	186,796	134,635	115,099	**2,806,961**
2007	259,014	251,012	257,060	257,344	270,208	273,754	283,386	274,043	206,210	183,758	134,211	117,326	**2,767,326**
2008	250,838	255,283	250,670	254,673	263,320	270,070	273,465	272,087	204,451	182,259	131,089	119,287	**2,727,492**
2009	250,063	250,786	257,512	250,275	263,672	265,493	273,498	263,266	209,370	184,151	130,120	116,538	**2,714,744**
2010	241,873	243,868	248,213	252,956	255,064	261,851	262,836	263,011	203,483	183,882	130,272	117,467	**2,664,776**
2011	238,385	237,493	243,691	243,155	259,063	254,334	261,481	251,302	200,526	176,525	130,773	117,429	**2,614,157**
2012	237,651	234,244	237,672	239,177	250,771	258,154	252,956	251,540	201,319	173,949	126,321	112,246	**2,576,000**
Total	**2,521,596**	**2,537,391**	**2,568,017**	**2,600,096**	**2,706,176**	**2,746,702**	**2,787,081**	**2,721,545**	**2,031,083**	**1,794,467**	**1,291,849**	**1,122,803**	**27,428,806**

Notes: **(i)** This table correspond to our student panel dataset (SPD), where every student has associated a unique observation in every year. **(ii)** Each student cohorts is followed in the downward sloping diagonal. For example, pupils in 1st grade in 2003 should be attending 4th grade in 2006 and 10th grade in 2012. **(iii)** After 8th grade the number of students observed drops dramatically because we do not follow students in technical schools.

Breaking down the SPD, and separating the data by school type of dependence we can see the distribution of students by type of schools during the period 2003-2012.

As we have mentioned in the Introduction, there are three types of school dependence in the Chilean education system: (i) *Municipal* schools; (ii) *Private Voucher* schools; and (iii) *Unsubsidised Private* schools.

Along the panel, we observe that *Municipal* schools have decreased the number of students enrolled, reducing in approximately 10% its share. The natural transition seems to be from *Municipal* to *Private Voucher* schools, while *Unsubsidised Private* schools maintain relatively stable the number of students enrolled.

Table 3.7: Number and distribution of students by type of dependence

Year	School Dependence			Total
	Municipal	Priv. Voucher	Unsub. Priv.	
2003	1,499,831	1,138,998	244,260	2,883,089
2004	1,450,799	1,153,205	241,049	2,845,053
2005	1,411,150	1,203,455	214,475	2,829,080
2006	1,350,089	1,246,200	210,672	2,806,961
2007	1,289,377	1,264,283	213,666	2,767,326
2008	1,222,237	1,290,551	214,704	2,727,492
2009	1,177,515	1,321,748	215,481	2,714,744
2010	1,112,082	1,336,357	216,337	2,664,776
2011	1,060,762	1,336,557	216,838	2,614,157
2012	1,010,626	1,350,974	214,400	2,576,000
Total	**12,584,468**	**12,642,328**	**2,201,882**	**27,428,678**
Average	**1,343,000**	**1,231,206**	**222,044**	**2,796,249**

Year	School Dependence (%)			Total
	Municipal	Priv. Voucher	Unsub. Priv.	
2003	52.0%	39.5%	8.5%	100%
2004	51.0%	40.5%	8.5%	100%
2005	49.9%	42.5%	7.6%	100%
2006	48.1%	44.4%	7.5%	100%
2007	46.6%	45.7%	7.7%	100%
2008	44.8%	47.3%	7.9%	100%
2009	43.4%	48.7%	7.9%	100%
2010	41.7%	50.1%	8.1%	100%
2011	40.6%	51.1%	8.3%	100%
2012	39.2%	52.4%	8.3%	100%
Total	**46%**	**46%**	**8%**	**100%**
Average	**46%**	**46%**	**8%**	**100%**

Notes: **(i)** The type of school dependence refers to the schools ownership and the type of funding. **(ii)** *Private Voucher* (Priv. Voucher) schools get public funds from pupils voucher, while *Unsubsidised Private* (Unsub. Priv) schools exclusively privately founded.

In terms of schools, from Table 3.8, we see a decrease in the number of *Municipal* schools serving students, where more than 10% of schools observed in 2003 are not longer in the panel by 2012. The observed reduction could be explained by changes in the type of school dependence or just because of schools shutting down. The *Unsubsidised private schools* also experienced a reduction in the number of schools, but what it is most likely happening here is that many of them turnout and became *Private Voucher* schools.

Table 3.8: Number and distribution of schools by type of dependence

Year	School Dependence			Total
	Municipal	Priv. Voucher	Unsub. Priv.	
2003	5,715	3,027	582	9,324
2004	5,673	3,051	559	9,283
2005	5,658	3,209	462	9,329
2006	5,530	3,321	448	9,299
2007	5,462	3,363	448	9,273
2008	5,397	3,425	447	9,269
2009	5,367	3,500	440	9,307
2010	5,290	3,513	443	9,246
2011	5,146	3,533	436	9,115
2012	5,074	3,573	431	9,078
Average	**5,543**	**3,271**	**484**	**9,298**

Year	School Dependence (%)			Total
	Municipal	Priv. Voucher	Unsub. Priv.	
2003	61.3%	32.5%	6.2%	100%
2004	61.1%	32.9%	6.0%	100%
2005	60.6%	34.4%	5.0%	100%
2006	59.5%	35.7%	4.8%	100%
2007	58.9%	36.3%	4.8%	100%
2008	58.2%	37.0%	4.8%	100%
2009	57.7%	37.6%	4.7%	100%
2010	57.2%	38.0%	4.8%	100%
2011	56.5%	38.8%	4.8%	100%
2012	55.9%	39.4%	4.7%	100%
Average	**60%**	**35%**	**8%**	**100%**

Notes: (i) The type of school dependence refers to the schools ownership and the type of funding. **(ii)** *Private Voucher* (Priv. Voucher) schools get public funds from pupils voucher, while *Unsubsidised Private* (Unsub. Priv) schools exclusively privately founded.

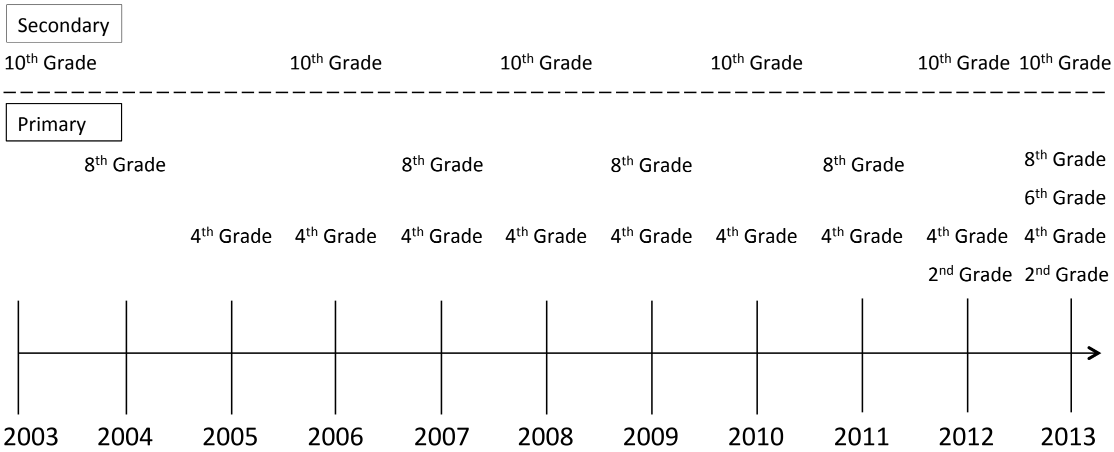
Notes: **(i)** The type of school dependence refers to the schools ownership and the type of funding. **(ii)** *Private Voucher* (Priv. Voucher) schools get public funds from pupils voucher, while *Unsubsidised Private* (Unsub. Priv) schools exclusively privately founded.

3.5 Selected cohorts for correlation analysis

In the following sections we present an exhaustive analysis of the correlation between standardised Simce scores and school marks standardised at school-grade level. Knowing that the National Examination is not taken in all grades every year, we want to to find out an alternative to replace the Simce scores when two consecutive measures of academic performance are required.

Figure 3.2 presents in which grades the Simce exam has been taken during the period 2003 - 2013. The Simce has been mainly taken in 4^{th}, 8^{th} and 10^{th} grades, where 8^{th} and 10^{th} were tested every other year. Just from 2013, the National examination started to be taken every year in 4^{th}, 6^{th}, 8^{th} and 10^{th} grades. In addition, 2^{nd} grade which started being taken, but only in Language.

Figure 3.2: Timeline - Simce



Note: The grades we show in this timeline are those where Simce is taken for the corresponding years. We separate between Primary and Secondary grades.
Source: The National Examination (Simce) data set.

The results obtained from the Simce exam can be used as measures of pupil cognitive abilities, which is what we need for modelling Value Added measures. Pupil academic performance, represented by Simce scores can be explained by unobserved heterogeneities such as teacher and school effects. However, Value Added Models require consecutive measures of pupil achievement to capture the impacts of teachers (and schools) from one year to the next. From Figure 3.2 we can see that cohorts taking the Simce exam in 4^{th} grade do not take the exam either in 3^{rd} or 5^{th} grade, which it would be desirable for most of the Value Added Models.

To address the lack of Simce scores for consecutive grades, we propose to use school marks standardised at school-grade level. Language and Maths school marks correspond to the final grades obtained for each pupil in every subject, and they assess the same pupil cognitive abilities we require for academic measures in the Value Added Models.

Although we are aware of the difference of evaluation criteria between school marking and National Examination testing, we investigate whether standardised Language and Maths marks at school level can be used as good proxies for standardised Language and Maths Simce scores. Firstly, we analyse the correlation levels for three different 4^{th} grade cohorts. Secondly, we compare graphically kernel distributions of both academic measures, and finally we apply a regression analysis where we confirm the positive correlation between standardised school marks and standardised Simce scores.

We select three 4^{th} grade cohorts to carry out our analyses. The selected cohorts we show in Figure 3.3 are representative of the five 4^{th} grade cohorts, from 2005 to 2009, which we will use in this thesis.

The selected cohorts are taken from the SPD, which is considered our master base, and we merge the Simce scores (from the National Examination data base) and the Language and Maths marks (from the School Marks data base) for our correlation analysis. The key matching variable between the databases is the unique student identification number (Mrun).

Figure 3.3: Selected cohorts

	2004	2005	2006	2007	2008	2009
4^{th} Grade		Cohort 1		Cohort 2		Cohort 3

Note: Cohort 1(a) and Cohort 1(b) are mainly composed by the same group of students which have passed from 4th grade to 8th grade without repeating. The differences between students could be explained by other students who repeated at least once during this period, plus other attrition problems.

3.5.1 Student panel and performance measure matching

The matching process of the SPD to the National Examination and School Marks DBs generates some missing observations. However, we have created a register for those cases where Simce scores and school marks are not available, and we will use them for selection purposes when we estimate our Value Added Models. In Table 3.9 we show the matching between the the two performance data bases (Simce and School Marks), where we take the Simce Scores (SScs), Language Marks (LMrk) and Maths Marks (MMrk) to be assigned to every pupil in the cohort. The percentage of missing values is presented in every case, and we can observe their availability across cohorts.

In all cohorts up to 2007, we can see how the rate of missing observations when matching to the individual Simce score stays around 7%, while in 2009 it increases to 11%, approximately. It is not clear why the rate of missing Simce Scores increased in 2009.

There are some reasons why a school, and therefore a student, does not have a Simce score. These reasons are related to the minimum number of students taking the exam and the absenteeism rate on the day of the exam. There is a list of requirements that schools have to fulfil. When any of requirements fails, the results of the exam are not provided. The requirements could vary from year to year depending on the design of the exam.

Table 3.9: Match between performance data bases (selected cohorts)

	Student panel: 4th Grade Cohorts			
2005	With Match	Without Match	Total	% Missing
Simce Score (SScs)	248,819	19,343	268,162	7.2%
Lang Marks (LMrk)	263,872	4,290	268,162	1.6%
Maths Marks (MMrk)	264,936	3,226	268,162	1.2%
Both SSmc - LMrk	246,854	2,325	249,179	0.9%
Both SSmc - MMrk	247,618	2,025	249,643	0.8%
All SSmc - LMrk - MMrk	246,498	2,009	248,507	0.8%
2007	With Match	Without Match	Total	% Missing
Simce Score (SScs)	238,785	18,559	257,344	7.2%
Lang Marks (LMrk)	253,033	4,311	257,344	1.7%
Maths Marks (MMrk)	253,972	3,372	257,344	1.3%
Both SSmc - LMrk	236,913	2,439	239,352	1.0%
Both SSmc - MMrk	237,554	2,141	239,695	0.9%
All SSmc - LMrk - MMrk	236,372	2,131	238,503	0.9%
2009	With Match	Without Match	Total	% Missing
Simce Score (SScs)	222,933	27,342	250,275	10.9%
Lang Marks (LMrk)	247,385	2,890	250,275	1.2%
Maths Marks (MMrk)	248,424	1,851	250,275	0.7%
Both SSmc - LMrk	222,075	2,310	224,385	1.0%
Both SSmc - MMrk	222,807	1,725	224,532	0.8%
All SSmc - LMrk - MMrk	222,075	1,725	223,800	0.8%

Note: In this table we represent the matching observed between the Student Panel Dataset (SPD), Simce scores (SScs), Language marks (LMrk) and Maths Marks (MMrk), for every selected cohort.

3.6 Correlation evidence

The correlation analysis presented in this section is composed by matrices and graphs correlations, kernel distributions, and linear regressions.

For every selected 4^{th} grade cohort, we check the correlation levels of the Language and Maths marks with their respective scores in the National exam. Furthermore, we disaggregate the correlation analysis by type of school dependence, identifying possible differences in marking distributions between *Municipal*, *Private Voucher* and *Unsubsidised Private* schools.

In order to make these performance measures comparable, we have standardised the Simce scores at national level, and the school marks at the school-grade level. Our hypothesis is that a good student within a school as measured by school marks, would be a good predictor for his or her performance as measured by the National Examination Simce score.

3.6.1 Correlation matrices

We present the correlation matrices for standardised school marks and the standardised Simce scores in the selected cohorts. Each matrix considers marks and scores per subject. Then we conduct the correlation analysis by subject.

Table 3.10 shows a very stable correlation between school marks and Simce scores in all cohorts, although being slightly higher in Maths. In Table 3.11, we present the correlation matrices separated by type of schools dependence. The correlation level increases to approximately 0.61 to 0.64 in *Municipal* schools, but it keeps similar in *Private Voucher* and *Unsubsidised Private* schools, between 0.56 to 0.62 approximately. Regarding rurality conditions of school, we can see from Table 3.12 that correlation keeps similar and stable in urban and rural areas.

Table 3.10: General correlations - Selected cohorts

	Cohort 1 - 4th Grade 2005			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.56**	1		
Stdz Maths School Marks	0.82	0.51	1	
Stdz Maths Simce Score	0.54	0.78	**0.59**	1

Observations: 240,812

	Cohort 2 - 4th Grade 2007			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.57**	1		
Stdz Maths School Marks	0.80	0.50	1	
Stdz Maths Simce Score	0.53	0.78	**0.58**	1

Observations: 230,096

	Cohort 3 - 4th Grade 2009			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.56**	1		
Stdz Maths School Marks	0.79	0.48	1	
Stdz Maths Simce Score	0.52	0.75	**0.58**	1

Observations: 211,960

Table 3.11: Correlation matrices by type of school dependence - 4^{th} grade cohorts

	Cohort 1 - 4th Grade 2005 Municipal Schools			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.62**	1		
Stdz Maths School Marks	0.83	0.56	1	
Stdz Maths Simce Score	0.59	0.77	**0.64**	1

Observations: 120,379

	Cohort 2 - 4th Grade 2007 Municipal Schools			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.61**	1		
Stdz Maths School Marks	0.81	0.54	1	
Stdz Maths Simce Score	0.59	0.76	**0.63**	1

Observations: 107,885

	Cohort 3 - 4th Grade 2009 Municipal Schools			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.60**	1		
Stdz Maths School Marks	0.79	0.52	1	
Stdz Maths Simce Score	0.58	0.73	**0.63**	1

Observations: 87,430

	Cohort 1 - 4th Grade 2005 Private Voucher Schools			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.57**	1		
Stdz Maths School Marks	0.81	0.51	1	
Stdz Maths Simce Score	0.54	0.77	**0.60**	1

Observations: 105,652

	Cohort 2 - 4th Grade 2007 Private Voucher Schools			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.59**	1		
Stdz Maths School Marks	0.80	0.52	1	
Stdz Maths Simce Score	0.55	0.76	**0.60**	1

Observations: 107,572

	Cohort 3 - 4th Grade 2009 Private Voucher Schools			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.58**	1		
Stdz Maths School Marks	0.79	0.49	1	
Stdz Maths Simce Score	0.55	0.74	**0.61**	1

Observations: 109,649

	Cohort 1 - 4th Grade 2005 Unsubsidised Private School			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.55**	1		
Stdz Maths School Marks	0.78	0.50	1	
Stdz Maths Simce Score	0.51	0.65	**0.60**	1

Observations: 14,781

	Cohort 2 - 4th Grade 2007 Unsubsidised Private School			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.59**	1		
Stdz Maths School Marks	0.77	0.51	1	
Stdz Maths Simce Score	0.51	0.67	**0.61**	1

Observations: 14,639

	Cohort 3 - 4th Grade 2009 Unsubsidised Private School			
	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.57**	1		
Stdz Maths School Marks	0.76	0.48	1	
Stdz Maths Simce Score	0.55	0.68	**0.62**	1

Observations: 14,881

Table 3.12: Correlation matrices by rurality - 4^{th} grade cohorts

4th Grade 2005 — Urban

	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	0.57	1		
Stdz Maths School Marks	0.81	0.51	1	
Stdz Maths Simce Score	0.54	0.78	0.59	1

Observations: 211,326

4th Grade 2007 — Urban

	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.57**	1		
Stdz Maths School Marks	0.80	0.51	1	
Stdz Maths Simce Score	0.53	0.78	**0.58**	1

Observations: 202,448

4th Grade 2009 — Urban

	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.56**	1		
Stdz Maths School Marks	0.79	0.48	1	
Stdz Maths Simce Score	0.53	0.76	**0.58**	1

Observations: 187,320

4th Grade 2005 — Rural

	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	0.58	1		
Stdz Maths School Marks	0.82	0.53	1	
Stdz Maths Simce Score	0.56	0.76	0.61	1

Observations: 29,486

4th Grade 2007 — Rural

	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.58**	1		
Stdz Maths School Marks	0.81	0.52	1	
Stdz Maths Simce Score	0.56	0.75	**0.60**	1

Observations: 27,648

4th Grade 2009 — Rural

	Stdz Lang School Marks	Stdz Lang Simce Score	Stdz Maths School Marks	Stdz Maths Simce Score
Stdz Lang School Marks	1			
Stdz Lang Simce Score	**0.57**	1		
Stdz Maths School Marks	0.80	0.50	1	
Stdz Maths Simce Score	0.53	0.72	**0.58**	1

Observations: 24,640

The differences observed in correlation levels for *Unsubsidised Private* schools could suggest uneven distributions in the academic performance between type of school dependence. The following section shows kernel density graphs as measurement instrument to compare distributions between standardised school marks and standardised Simce scores.[9]

3.6.2 Kernel distributions

The kernel function is used to estimate the density of our variables of interest; standardised school marks and standardised Simce scores. The analysis consists in analyse how similar the aggregated distribution per subject is, and how the differences are when we disaggregate standardised school marks and standardised Simce scores by type of school dependence.

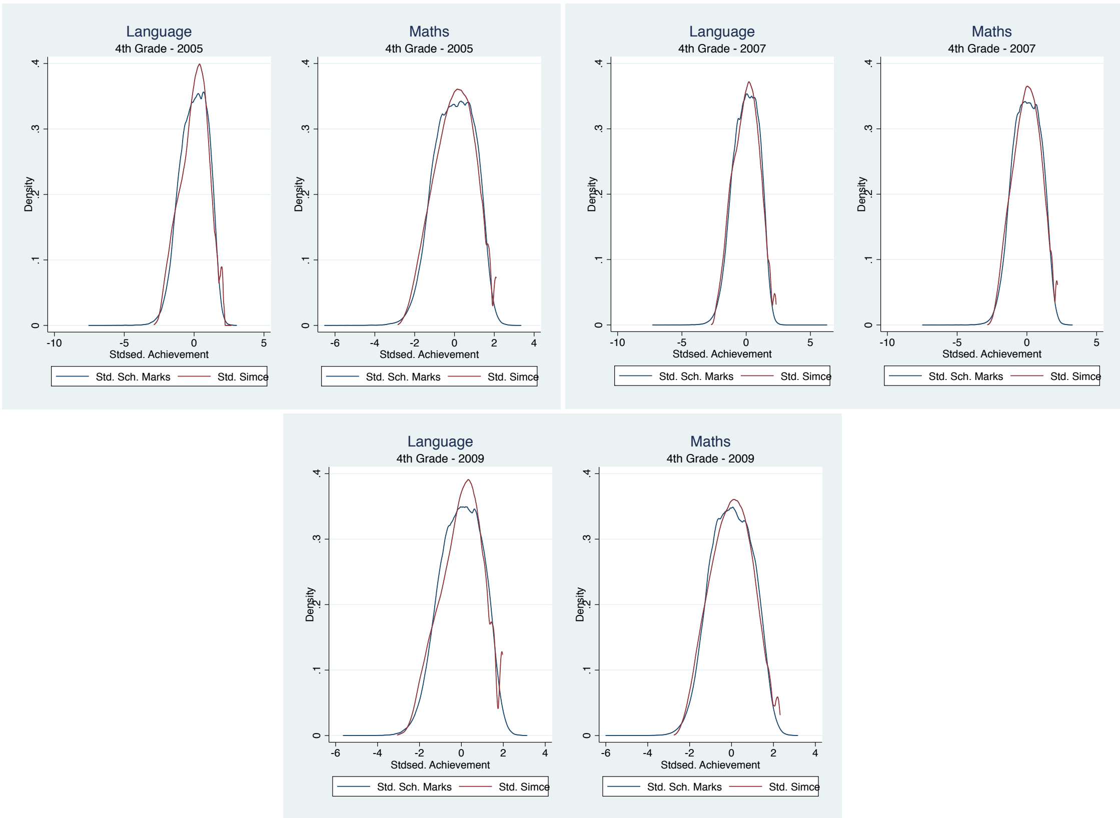
School Marks vs Simce Scores - Aggregate level

Comparing the distributions of standardised school marks and standardised Simce scores, it is possible to confirm they both behave similarly in aggregated terms. In the figures below we present the comparison of the distributions by subjects for the three selected 4^{th} grade cohorts.

The shapes of the distributions differ only a little between Language and Maths standardised measures. Along the three selected cohort the distributions and the differences between subjects are very similar.

[9] In Appendix 3.5, we present scatter plots for correlation levels in every cohort.

Figure 3.4: Stdsed. School Marks vs Stdsed. Simce Scores
4^{th} grade cohorts 2005-2007-2009



In the following subsection, we look for some evidence of what could explain differences in the kernel distributions of school marks and Simce exam results, by type of school dependence.

School Marks vs Simce Scores - Disaggregated by type of school

Separating the analysis by type of school dependence, it is easy to recognise some clearer differences in marks distributions ruled by *Unsubsidised Private* schools. The difference is very visible when comparing the standardised results for the National Examination.

Differences in distributions for *Municipal* and *Private Voucher* schools are not important, as it can be seen from Figures 3.5 and 3.6.

On the other hand, the plots in Figure 3.7 confirm that Simce scores in the *Unsubsidised Private* schools are more concentrated to the right-hand side of the distribution mean. This confirm that among the total population, pupils from *Unsubsidised Private* school perform better in the Simce exam, and the distribution of standardised school marks might underestimate their academic achievement in terms of Simce scores. Although, the share of students enrolled in *Unsubsidised Private* school is less than 10% of the total, the differences observed in the standardised National scores could drive the small differences found in the comparison of the aggregate distributions above.

Figure 3.5: Kernel distributions in Municipal schools
4^{th} grade cohorts 2005-2007-2009

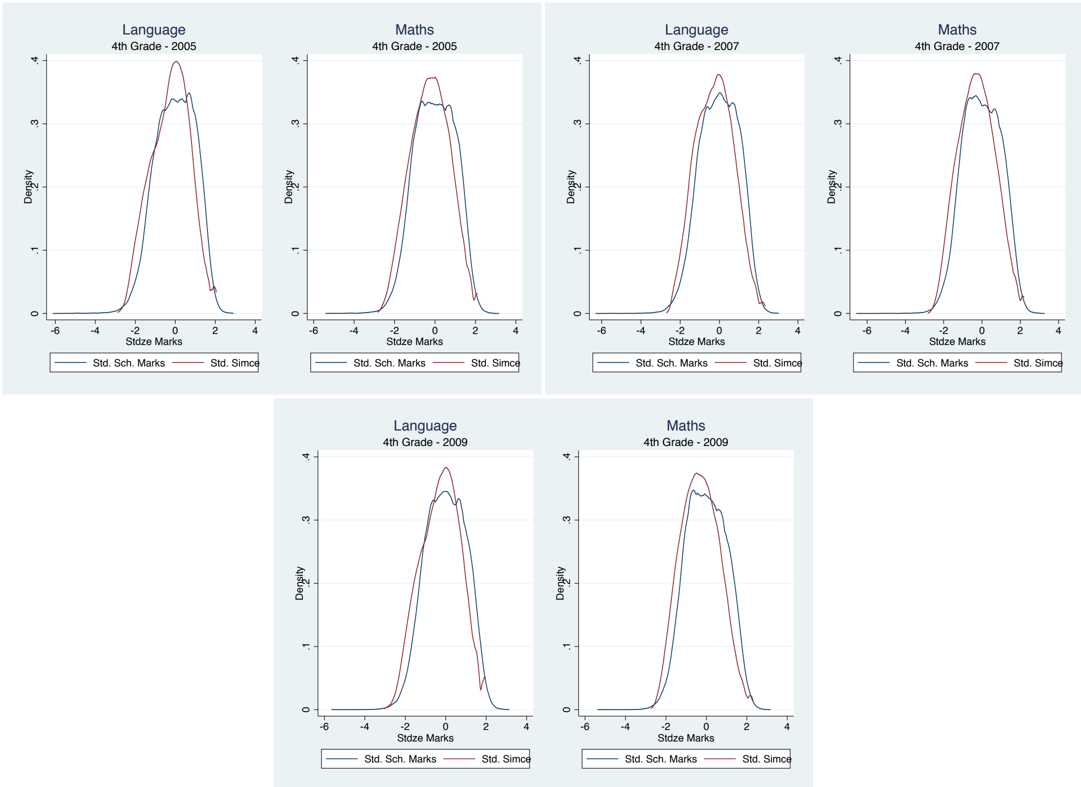


Figure 3.6: Kernel distributions in Private Voucher schools
4^{th} grade cohorts 2005-2007-2009

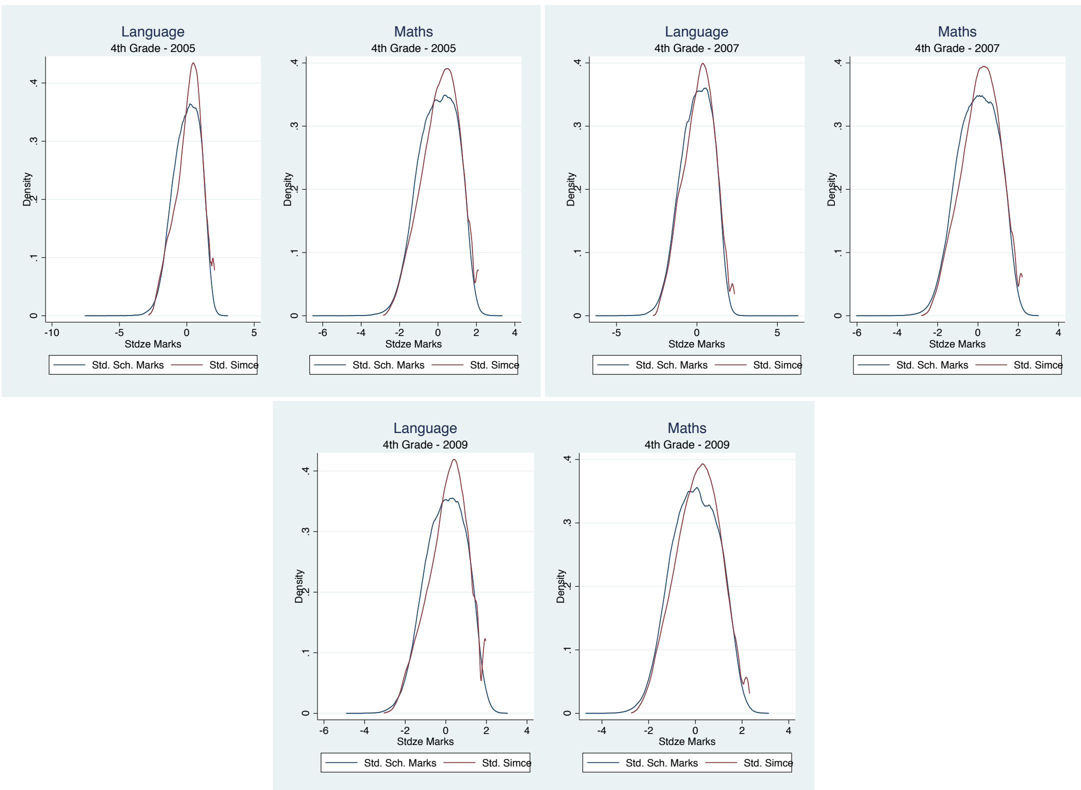
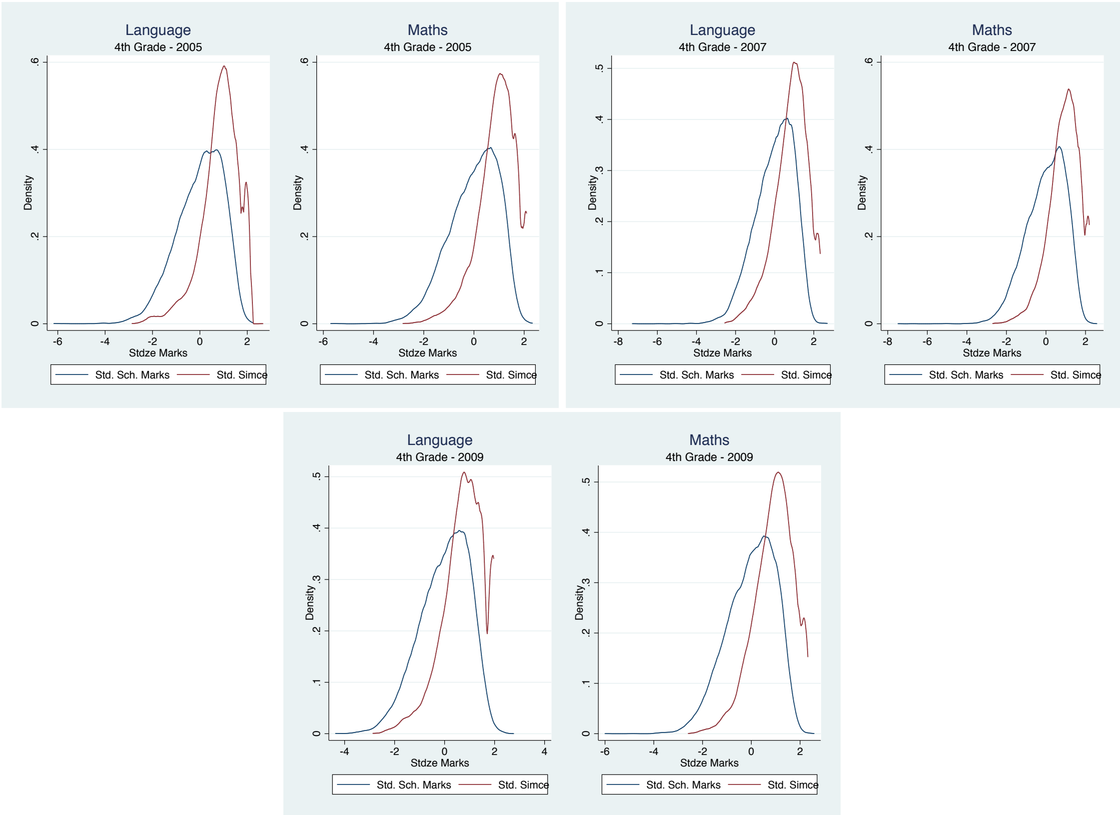


Figure 3.7: Kernel distributions in Unsubsidised Private schools
4^{th} grade cohorts 2005-2007-2009



Even though we are not able to claim quality differences between type of schools based on correlation analysis, we suggest controlling for the type of school when further estimations will be done.

3.6.3 Regression analysis

Here we add the third piece of evidence indicating a positive correlation between standardised school marks and National Simce scores. This analysis consists in estimating single linear regressions of standardised Language and Maths Simce score on their respective standardised school marks by subject.

The estimation results suggest that correlation between both performance measures is statistically significant. Table 3.13 show the results obtained for all three 4^{th} grade selected cohorts.

The estimated coefficients on the standardised Language and Maths marks are between 0.58 and 0.61 in all three 4^{th} grade cohorts. The differences between subject are hold across cohorts. However, all estimates are significant at 1% of significance level.

When we add controls for type of school dependence, the correlation increases slightly holding the significance level at 1% in both cohorts.

Table 3.13: Linear regressions
4^{th} grade cohorts 2005-2007-2009

	Cohort 1 - 2005				Cohort 2 - 2007				Cohort 3 - 2009			
	Stdz Simce - Lang	Stdz Simce - Maths	Stdz Simce - Lang	Stdz Simce - Maths	Stdz Simce - Lang	Stdz Simce - Maths	Stdz Simce - Lang	Stdz Simce - Maths	Stdz Simce - Lang	Stdz Simce - Maths	Stdz Simce - Lang	Stdz Simce - Maths
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Stdz Language school marks	0.587***		0.592***		0.594***		0.599***		0.579***		0.586***	
	(0.002)		(0.002)		(0.002)		(0.002)		(0.002)		(0.002)	
Stdz Maths school marks		0.610***		0.614***		0.598***	(0.007)	(0.006)		0.595***		0.604***
		(0.002)		(0.002)		(0.002)		0.602***		(0.002)		(0.002)
Controls								(0.002)				
Types of school dependence	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Constant	-0.032***	-0.030***	-0.273***	-0.270***	-0.033***	-0.029***	-0.290***	-0.311***	-0.040***	-0.036***	-0.314***	-0.358***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Observations	243,874	244,499	243,874	244,499	233,319	234,267	233,319	234,267	216,693	217,296	216,693	217,296
R-squared	0.319	0.347	0.403	0.438	0.327	0.334	0.413	0.439	0.311	0.331	0.390	0.445

Notes: (i) Type of school dependence controls refer to dummy variables in relation to Municipal, Private Voucher and Unsubsidised Private schools, where we leave Municipal schools as a pivot dummy variable. **(ii)** *** p<0.01; ** p <0.05; * p <0.1

These results confirm the positive relationships observed in the previous analyses. Graphically and statistically the correlation between standardised school marks and standardised Simce scores is positive, relatively high and significant.

3.7 Conclusion

We have described in some detail the composition of the available data sets that will provide the basis for our empirical analysis, pointing at both potential strengths and limitations of the data.

First, we construct our own Student Panel Dataset (SPD) from the administrative data, and we describe how it can be enriched with other complementary data sets, such as the National Examination, School Marks and Teachers data bases.

Regarding the limitations of the data, the main restriction is the lack of National Examination score for every grade in every year. That is why the challenge is to find a consistent substitute for individual performance achievement when one is required but the Simce scores are not available or not observed.

Throughout the second part of this chapter we study the correlation between standardised school marks and Simce scores per subject. The analysis is based on three selected 4^{th} cohorts, for representative years in our analyses (2005, 2007 and 2009).

Three different methodologies are applied to the correlation analysis. For all methodologies we work with standardised values of school marks and of Simce scores. School marks are standardised at school-grade-year level and Simce scores are standardised at grade-year level. We first show simple correlation tables, then we present graphical distribution comparisons and lastly we run some linear regressions of standardised Simce scores on standardised school marks.

Correlation levels in 4^{th} grade cohorts are between 0.56 to 0.59. The correlation of standardised Maths school marks and standardised Maths Simce scores is slightly higher than Language correlation for all cohorts. When we disaggregate by type of school dependence the results do not vary substantially.

The graphical kernel distribution analysis did not show considerable differences when we compare standardised school marks and Simce scores by subject. However, when we disaggregate by type of school dependence and we compare between them, differences in distributions arise specially for the *Unsubsidised private* school. This type of school seems to drive any differences between performance measure distributions.

In the final part of the correlation analysis, we compare simple linear regressions of standardised Simce scores on standardised school marks, with and without type of school dependence controls. Results show that correlations slightly increase when we control for type of school dependence, but in both cases (with and without controls) the coefficients are significant at 1% of significance level.

Supported by three different methodologies, we suggest there is evidence that standardised school marks and standardised Simce scores are reasonably highly correlated. On this basis, we propose the use of standardised schools marks as proxies for standardised Simce scores in specific cases when the latter are not available.

Chapter 4

Testing non-random assignment: Evidence from the Chilean school system

Abstract

In this chapter we test for students-to-teacher non-random assignment within schools in the Chilean school system. We focus on possible student grouping in 4^{th} and 5^{th} grade classes based on previous academic achievement. To analyse non-random assignment within schools, we select a group of schools and classes where we could actually test the evidence of student sorting. We applied some selection criteria at students and class level to get a sample of all schools with 2 and 3 classes per grade.

We use graphical and statistical tests of evidence of student sorting. Using graphical analysis we were not able to identify any evidence of non-random assignment in the Chilean school system. We generate two types of counterfactual classes: (i) based on random student assignment, and (ii) based on perfectly sorted student assignment. We then compare both counterfactuals with real class distributions. To make these comparisons, we create five categories of non-random assignment evidence: *None, Low, Med-High* and *High*. All tested schools from the sample are classified in one of these categories.

Comparisons with respect to random counterfactuals do not show significant evidence of non-random assignment when we use graphical and statistical testing. However, when we compare real class distributions with perfectly sorted counterfactuals, the number of schools found with *Low* and *Medium* levels of sorting evidence increase, but not in higher ones. Our results suggest that most of the schools are concentrated in *None* or *Low* categories of non-random assignment evidence.

Contents

4.1 Introduction

One of the main concerns for the estimation of "Teacher Effects" (TEs) is the difficulty of interpreting estimated effects as causal on student outcomes. As we have discussed in Chapter 2, the Value Added Models (VAMs) are commonly used in the literature to estimate TEs relying on non-random assignment of students to teachers. Therefore, it is important to study whether in the Chilean school system context the non-random assignment assumptions are likely to hold.

In Chapter 2, we mentioned three different sources of non-random assignment which might cause VAM assumptions to fail. The cases of non-random assignment are: (1) **student-to-school**; (2) **teacher-to-school**, and (3) **student-to-teacher**.

The first source of non-random assignment is due to the free school choice system where parents are able to choose which school to apply for their children. *Private Voucher* and *Unsubsidised Private* schools can decide which students to accept, while *Municipal* schools, in theory, are not allowed to select. Regarding the second source of non-random assignment, the teachers to school allocation might be affected by selection issues, as the decision for hiring a teacher or accepting a job offer is taken endogenously. However, independent of the flexibility levels in terms of pupil selection or the teacher hiring process across schools, both types of assignments (**student-to-school** and **teacher-to-school**) can be modelled by a "Two-side Matching Model" which is currently beyoned our research agenda.

We follow the literature when TEs are estimated in non-experimental contexts. Thus, we rely on the **student-to-school** and **teacher-to-school** random assignment assumptions, controlling for as many as posible variables which might be the source of non-random assignment based on observed characteristics. However, mindful of the differences between type of schools, we control for types of school dependence, and we conduct throughout the thesis some separate analyses by type of schools.

We are focus on verifying **Student-to-teacher** (or **teachers-to-classrooms**) random assignment within schools. This type of assignment depends mainly on the principal's discretion, who based on observable or unobservable factors, may decide which grade and classroom a particular teacher has to teach. Moreover, principals can decide whether a primary teacher be assigned the role of *general teacher* or *subject specialist* (SS) teacher. When teachers are specialised in one specific subject, teaching it to more than one class, we refer to schools being under SS teacher scheme. It is very unlikely that principals make a random decision every year about teacher-classroom allocation.

Therefore, the third source of non-random assignment within school could happen when principals track students marks records and group pupils based on their previous performance. Principals could also sort students in terms of non-observable characteristics, but we just focus on ruling out the sorting evidence in terms of observable academic achievement.

We also explained in Chapter 2, that to estimate teacher and school effects we need that unobserved time-varying factors are not correlated with all the regressors included in the VAM. If students are not sorted into classrooms within school-grade, we can argue that teacher to classroom allocation is not correlated to observables or unobservable pupil achievement factors.

In this chapter we examine whether there is evidence of sorting in the Chilean school system considering selected samples of three different cohorts. Specifically, we analyse reduced panels looking for evidence of grouping based on previous school marks. Following Aaronson et al. (2007), our strategy consists of graphic analysis and statistical tests for measuring differences in means and distributions between real classes and two generated counterfactual classes. The two counterfactual distributions are: (i) "*Random Sorting*" counterfactuals which are constructed randomly, and (ii) "*Perfect Sorting*" counterfactual groups which are created based on marks ranking within school-grades. The *perfect sorted* counterfactuals are separated into *High* and *Low* performing students.

We replicate Aaronson et al. (2007) methodology in our selected cohorts, and we obtain results consistent with most weak evidence of non-random assignment. These results support our hypothesis that in the Chilean school system there is not significant evidence of non-random assignment of student-to-teacher between 4^{th} and 5^{th} grades, at least for schools with 2 and 3 classes per grade.

This chapter is organised as follows. In Section 1, we select three cohorts and generate reduced panels for each. Section 2 describes and compares these cohorts, showing in detail the sample selection process applied to the 4^{th} Grade 2005 cohort when we create its respective reduced panel (from 3^{th} to 5^{th} grade). In Section 4, we show graphic evidence suggesting that the assumption of non-random assignment of student-to-classroom is not problematic. Section 5 presents statistical tests for differences in means and distributions between real and counterfactual classrooms per school. Finally, in Section 6 we summarise the results and present our main conclusions.

As complementary information, we show in the "Appendix - Chapter 4" the description and results of the two other cohorts tested: 4^{th} Grade 2007 and 4^{th} Grade 2009.

4.2 Selected cohorts and mini panels description

Initially, we have generated the Student Panel Dataset (SPD), which is an unbalanced panel with all students officially registered in the Chilean school system from 2003 to 2012. From the SPD we take three representative cohorts in which we examine whether there is evidence of non-random assignment in early primary grades.

The selection of the three cohorts is based on the availability of the National standardised exam scores, which are going to be used for the teacher and school effects estimation in later chapters alongside standardised school marks. As we have seen in Chapter 3, the Chilean National Examination (Simce) started being taken in 4^{th} grade every year from 2005. In the thesis, we consider for our estimations 4^{th} grade cohorts from 2005 to 2009. Therefore, we select a representative cohorts sample for this period choosing the following cohorts:

- **Cohort 1**: 4^{th} Grade 2005
- **Cohort 2**: 4^{th} Grade 2007
- **Cohort 3**: 4^{th} Grade 2009

From these three cohorts we want to track pupils who were in 3^{rd}, 4^{th} and 5^{th} grade in consecutive years without school changes. Identifying those cases would allow us to examine whether there is evidence of non-random assignment based on school marks within schools when 4^{th} and 5^{th} grade classes are arranged.[1] We claim that if there is an earlier pupil grouping it is reflected in this analysis.

For each selected cohort we generate a reduced 3 year panel (we also call it mini panel). We are particularly interested in students who followed the regular transition from 3^{rd} to 5^{th} grade (e.g. without repeating grades), considering as the reference group those enrolled in the 4^{th} grades for the chosen years.

[1] We do not propose that pupil sorting can be made based on Simce scores, as the Simce score per pupil is not available for principals or school authorities. Schools only observe the average performance of the cohort. Then, the only measure of pupil performance that the principal can use for grouping students into classes are the school marks.

Table 4.1: Reduced panels for selected cohorts

	Cohort 1				
	3rd Grade	4th Grade	5th Grade	Other/Miss	Total
2004	252,423	8,246	517	6,976	268,162
2005		**268,162**			268,162
2006	65	8,589	253,419	6,089	268,162
	Cohort 2				
	3rd Grade	4th Grade	5th Grade	Other/Miss	Total
2006	243,337	8,944	63	5,000	257,344
2007		**257,344**			257,344
2008	31	8,020	244,683	4,610	257,344
	Cohort 3				
	3rd Grade	4th Grade	5th Grade	Other/Miss	Total
2008	237,417	7,657	8	5,193	250,275
2009		**250,275**			250,275
2010	1	6,636	240,986	2,652	250,275

Note: Using enrolled students in 4th grade as reference for every cohort, we see where they were one year before and one year ahead. In the first row of every table we identify in 4th grade the repeating students, and analogous for the third row, we found the students who were not promoted to 5th grade. The 5th grade in the first row and the 3rd grade in the third are cases which we assume as typo. In the Other/Miss column we add the rest of the typos in terms of grades and missing cases as these pupils were not observed in that year.

We describe the transitions of selected cohorts in Table 4.1, where we also identify missing observations, students repeating grades, and pupils who appear enrolled in infeasible grades (e.g. If a student from 4^{th} grade 2005 appears to be enrolled in 5^{th} grade in 2004 or in 3^{rd} in 2006).

Table 4.2: Selected cohorts transition

	Cohort 1				Cohort 2				Cohort 3		
In 2004	Fequency	Percent.	Cumulate	**In 2006**	Fequency	Percent.	Cumulate	**In 2008**	Fequency	Percent.	Cumulate
3th Grade	252,423	94%	94%	3th Grade	243,337	95%	95%	3th Grade	237,417	95%	95%
4th Grade	8,246	3%	97%	4th Grade	8,944	3%	98%	4th Grade	7,657	3%	98%
Other grade	1,669	1%	98%	Other grade	735	0%	98%	Other grade	401	0%	98%
Missing	5,824	2%	100%	Missing	4,328	2%	100%	Missing	4,800	2%	100%
Total	268,162	100%	100%	Total	257,344	100%	100%	Total	250,275	100%	100%
% Other/Miss	*2.8%*			*% Other/Miss*	*2.0%*			*% Other/Miss*	*2.1%*		
In 2006	Fequency	Percent.	Cumulate	**In 2008**	Fequency	Percent.	Cumulate	**In 2010**	Fequency	Percent.	Cumulate
4th Grade	8,589	3%	3%	4th Grade	8,020	3%	3%	4th Grade	6,636	3%	3%
5th Grade	253,419	95%	98%	5th Grade	244,683	95%	98%	5th Grade	240,986	96%	99%
Other grade	1,240	0%	98%	Other grade	749	0%	98%	Other grade	115	0%	99%
Missing	4,914	2%	100%	Missing	3,892	2%	100%	Missing	2,538	1%	100%
Total	268,162	100%	100%	Total	257,344	100%	100%	Total	250,275	100%	100%
% Other/Miss	*2.3%*			*% Other/Miss*	*1.8%*			*% Other/Miss*	*1.1%*		

Notes: (i) Other grade is when the student from the selected cohort is observed in a grade which does not correspond to the possible grades where pupil could be enrolled one year before or after she/he was enrolled in 4th grade. **(ii)** Missing is when the student from the selected cohort is not observed in the Student Panel one year before or after she/he was enrolled in 4th grade. **(iii)** The Other/Miss indicator corresponds to the percentage of pupils in other grades or without register for that particular year.

In Table 4.2 we present the cohort transitions in percentages. It is a complement of Table 4.1, and here we can see how approximately 95% of the students from the selected cohorts were found in their correspondent grade relative to the year. The rate of repetition in 4^{th} grade is stable and is around 3%, while the rate of missing and misallocated cases is around 2% per year.

From now on, we will be focused just on those students who did not repeat and were found in their respective grades across the mini panel period. Thus, the first group of dropped observations correspond approximately to 5% of the students from each original selected cohort.

4.2.1 Student panel dataset (SPD) and school marks matching

Considering only students who were observed continuously in 3^{rd}, 4^{th} and 5^{th} grades along the mini panel, we have merged their school marks for measuring non-random assignment. The matching process generates some missing observations, as some of the pupils did not appear in the School Mark Data Base. In Table 4.3 we describe the number of pupil observations with and without recorded GPA, Language and Maths marks.

The missing marks cases keep stable around 1% to 2%. Language marks present the highest missing percentage, followed by Maths marks. However, we have enough observations for assessing student grouping evidence based on previous marks.

Table 4.3: Matching school marks to students in selected cohorts

		3th Grade 2004						4th Grade 2005						5th Grade 2006					
		Lang Marks		Maths Marks		GPA		Lang Marks		Maths Marks		GPA		Lang Marks		Maths Marks		GPA	
Cohort 1	With Register	250,002	99%	251,156	99%	252,106	100%	263,872	98%	264,936	99%	267,596	100%	249,389	98%	250,441	99%	253,053	100%
	W/O Register	2,421	1%	1,267	1%	317	0%	4,290	2%	3,226	1%	566	0%	4,030	2%	2,978	1%	366	0%
	Total	252,423		252,423		252,423		268,162		268,162		268,162		253,419		253,419		253,419	
	% W/O Register	*1.0%*		*0.5%*		*0.1%*		*1.6%*		*1.2%*		*0.2%*		*1.6%*		*1.2%*		*0.1%*	

		3th Grade 2006						4th Grade 2007						5th Grade 2008					
		Lang Marks		Maths Marks		GPA		Lang Marks		Maths Marks		GPA		Lang Marks		Maths Marks		GPA	
Cohort 2	With Register	241,091	99%	241,992	99%	243,099	100%	253,033	98%	253,972	99%	256,547	100%	241,025	99%	242,276	99%	244,253	100%
	W/O Register	2,246	1%	1,345	1%	238	0%	4,311	2%	3,372	1%	797	0%	3,658	1%	2,407	1%	430	0%
	Total	243,337		243,337		243,337		257,344		257,344		257,344		244,683		244,683		244,683	
	% W/O Register	*0.9%*		*0.6%*		*0.1%*		*1.7%*		*1.3%*		*0.3%*		*1.5%*		*1.0%*		*0.2%*	

		3th Grade 2008						4th Grade 2009						5th Grade 2010					
		Lang Marks		Maths Marks		GPA		Lang Marks		Maths Marks		GPA		Lang Marks		Maths Marks		GPA	
Cohort 3	With Register	235,298	99%	236,250	100%	237,115	100%	247,385	99%	248,424	99%	250,275	100%	237,645	99%	239,349	99%	240,927	100%
	W/O Register	2,119	1%	1,167	0%	302	0%	2,890	1%	1,851	1%	0	0%	3,341	1%	1,637	1%	59	0%
	Total	237,417		237,417		237,417		250,275		250,275		250,275		240,986		240,986		240,986	
	% W/O Register	*0.9%*		*0.5%*		*0.1%*		*1.2%*		*0.7%*		*0.0%*		*1.4%*		*0.7%*		*0.0%*	

Notes: **(i)** With Register is when it is possible to match their GPA, Language and Maths school marks. **(ii)** The GPA comes from the Performance Data Base. The Language and Maths marks come from the Student Marks Data Base. Both data bases belong to the Administrative data set. See Chapter 2.

4.2.2 Student panel dataset (SPD) and teacher register matching

We separate the teacher dataset into Language and Maths teacher data bases and we merged them to the selected cohorts. In Table 4.4 we show there are some observations without teachers matching (Maths and Language teachers). When there is only one teacher identified per class, we automatically assign the teacher ID for the other subject which was not matched. In those cases where the SPD do not match with the Language and Maths teacher data set, we are not able to assign any teacher ID in these classes.

In our panel we classify teachers by subjects, although in early primary grades classes are taught mainly by *general teachers*. However, in later primary grades, from 5^{th} grade onwards, most of the teacher become *SS* teachers.

Describing the composition of students exposed to a *general* or a *SS* teacher, we can see Tables 4.5 and Table 4.6 how the number of classes taught by *SS* teachers is considerably higher in 5^{th} grade compared to 3^{rd} and 4^{th} grades.

Table 4.4: Matching teachers to classrooms by subject in selected cohorts

		3th Grade 2004						4th Grade 2005						5th Grade 2006					
		Lang Tch		Math Tch		Both		Lang Tch		Math Tch		Both		Lang Tch		Math Tch		Both	
Cohort 1	With Match	246,725	98%	247,892	98%	246,448	98%	262,473	98%	263,877	98%	261,894	98%	246,750	97%	248,362	98%	245,037	97%
	W/O Match	5,698	2%	4,531	2%	5,975	2%	5,689	2%	4,285	2%	6,268	2%	6,669	3%	5,057	2%	8,382	3%
	Total	252,423		252,423		252,423		268,162		268,162		268,162		253,419		253,419		253,419	
	% Non-Match	*2.3%*		*1.8%*		*2.4%*		*2.1%*		*1.6%*		*2.3%*		*2.6%*		*2.0%*		*3.3%*	
		3th Grade 2006						**4th Grade 2007**						**5th Grade 2008**					
		Lang Tch		Math Tch		Both		Lang Tch		Math Tch		Both		Lang Tch		Math Tch		Both	
Cohort 2	With Match	238,270	98%	239,166	98%	237,355	98%	252,233	98%	252,956	98%	250,925	98%	237,140	97%	238,620	98%	234,599	96%
	W/O Match	5,067	2%	4,171	2%	5,982	2%	5,111	2%	4,388	2%	6,419	2%	7,543	3%	6,063	2%	10,084	4%
	Total	243,337		243,337		243,337		257,344		257,344		257,344		244,683		244,683		244,683	
	% Non-Match	*2.1%*		*1.7%*		*2.5%*		*2.0%*		*1.7%*		*2.5%*		*3.1%*		*2.5%*		*4.1%*	
		3th Grade 2008						**4th Grade 2009**						**5th Grade 2010**					
		Lang Tch		Math Tch		Both		Lang Tch		Math Tch		Both		Lang Tch		Math Tch		Both	
Cohort 3	With Match	231,976	95%	232,966	96%	231,057	95%	247,523	96%	248,436	97%	246,965	96%	238,884	98%	240,235	98%	238,799	98%
	W/O Match	5,441	2%	4,451	2%	6,360	3%	2,752	1%	1,839	1%	3,310	1%	2,102	1%	751	0%	2,187	1%
	Total	237,417		237,417		237,417		250,275		250,275		250,275		240,986		240,986		240,986	
	% Non-Match	*2.3%*		*1.9%*		*2.7%*		*1.1%*		*0.7%*		*1.3%*		*0.9%*		*0.3%*		*0.9%*	

Notes: (i) Lang Tch: Language teachers; Maths Tch: Maths teachers. **(ii)** With Match is when the student observation from a specific school-classroom was successfully matched to the Language and Maths Teachers data bases. W/O Match is when the student observation from a specific school-classroom was unable to find any match with the Language and Maths Teachers data bases. **(iii)** The Both column indicates whether the Language and the Maths teacher simultaneously matched the student observation from a specific school-classroom. **(iv)** The Non-Match indicator corresponds to the percentage of individual observation from a particular school-classroom without Language teacher, Math teacher or both. **(v)** The Language and Maths Teacher data bases are generated form the Teachers Data Base which belongs to the Administrative data set (see Chapter 3).

Since 2004 to 2008 the number of 3^{rd} grade classes with *SS* teachers has grown by almost 100%. Similar rate of growth is experienced in 4^{th} grade classrooms from 2005 to 2009. The number of 4^{th} grade classes which changed from *General* to *SS* teacher also grew but in a much lower rate.

Table 4.5: Teacher classification at student level in selected cohorts

	3rd Grades			4th Grades			5th Grades		
	2004	2006	2008	2005	2007	2009	2006	2008	2010
General Teacher	221,342	203,321	186,235	219,008	196,608	172,001	34,134	28,744	22,481
Subject Specialist Teacher	25,106	34,034	44,822	42,886	54,317	74,964	210,903	205,855	216,318
Non-Match	5,975	5,982	6,360	6,268	6,419	3,310	8,382	10,084	2,187
Total	252,423	243,337	237,417	268,162	257,344	250,275	253,419	244,683	240,986
% Subject Specialist (SS)	*9.9%*	*14.0%*	*18.9%*	*16.0%*	*21.1%*	*30.0%*	*83.2%*	*84.1%*	*89.8%*

Note: Non-Match observations refer to pupils without neither Language nor Maths teacher register.

Table 4.6: Teacher classification at class level in selected cohorts

	3rd Grades			4th Grades			5th Grades		
	2004	2006	2008	2005	2007	2009	2006	2008	2010
General Teacher	10,254	9,584	8,994	9,689	8,840	8,066	3,410	3,085	2,682
Subject Specialist Teacher	993	1,324	1,700	1,589	1,998	2,717	7,525	7,452	7,911
Non-Match	246	237	246	236	240	145	300	363	113
Total	11,493	11,145	10,940	11,514	11,078	10,928	11,235	10,900	10,706
% Subject Specialist (SS)	*8.6%*	*11.9%*	*15.5%*	*13.8%*	*18.0%*	*24.9%*	*67.0%*	*68.4%*	*73.9%*

Note: Non-Match observations refer to pupils without neither Language nor Maths teacher register.

The significant increase in the number of classes taught by SS teachers in 3^{rd} and 4^{th} grades suggest changes in some academic strategies. We could infer that more principals are assigning SS teachers in early grades to prepare students for the Simce exam taken in 4^{th} grade.

In this section, we have eliminated all inconsistent pupil observations along the panel (e.g. those observations which do not even match with students repeating grades during the period). Additionally, we have identified all missing values generated when we merge the selected cohorts with the school marks and teacher data bases

The following section discusses the sample selection made on the 3^{rd}, 4^{th} and 5^{th} grade mini panel, with respect to 2004-2006 years. From the initial 268,162 students observed in 4^{th} 2005, we describe all selection process required to study non-random assignment in this group of students.[2]

4.3 Description of 4^{th} Grade 2005 mini panel and sample selection

To examine student potential sorting in 4^{th} and 5^{th} grades, we need to select a sample from the mini panels. The sample selection only considers schools with at least two classes per grade, and with students who were observed for three years in the same school.

In the current section we describe in detail the 4^{th} grade 2005 mini panel and its sample selection process. The descriptions of the other two 4^{th} grade cohorts are documented in Appendix 4.1. Before starting the sample selection from the mini panel, we present in Table 4.7 the descriptive statics of the main variables.

Grouping strategies can take place only in schools with at least two classes per grade. Schools with one class per grade cannot sort pupils by any observable or unobservable characteristic, as all of the enrolled students for a particular grade are in the same class. Of course, principals can still determine which teachers are

[2]The same analyses is carry out for **Cohorts 2** and **3**. The mini panel description tables are presented in Appendix 4.1.

assigned with each grade, but the non-random assignment due to the allocation of teachers to particular grades is not considered in this analysis.

To have an idea of how many schools could be grouping students non-randomly, we classify schools by the number of classes per grade. In Table 4.8, we show how in every year of the mini panel Cohort 1, the proportion of schools with only one class per grade is 73%. Thus, we will focus on the 27% left to apply the sample selection criteria.

Table 4.7: Summary statistics mini panel Cohort 1

Reference Cohort:			**N = 268,162**	
Period:			T = 3	
Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
Pupil Level				
GPA	5.68	0.74	1	7
Average school Language marks	5.42	0.77	1	7
Average school Maths marks	5.32	0.82	1	7
Gender (Female=1)	0.49	0.50	0	1
Age	9.77	0.75	7	15
Attendance	92.41	10.51	0	100
Special Needs	0.02	0.11	0	1
Class Level				
Available Language teacher	0.98	0.14	0	1
Available Maths teacher	0.98	0.12	0	1
No. Schools per Language teacher	1.05	0.21	1	3
No. Schools per Maths teacher	1.05	0.22	1	3
No. Grades per Language teacher	2.47	1.72	1	9
No. Grades per Maths teacher	2.52	1.74	1	8
No. Classes per Language teacher	2.47	1.72	1	9
No. Classes per Maths teacher	2.86	1.94	1	12
Subject Specialist teacher	0.30	0.46	0	1
School Level				
Municipal schools	0.60	0.49	0	1
Private Voucher schools	0.35	0.47	0	1
Unsubsidised Private schools	0.06	0.22	0	1
Rural Area	0.49	0.50	0	1
School Socioeconomic Level	1.26	1.19	0	4
Number of students per grade	30.76	36.55	1	525
Number of students per class	18.32	13.99	1	50
Number of students	268,162			
Number of classes	34,242			
Number of schools	8,681			

Notes: **(i)** The number of students corresponds to the total pupil observed from the original cohort in 4th grade 2005, and who were track across the three year period. **(ii)** The number of classes corresponds to all the classrooms observed over the three year periods in all schools. **(iii)** The number of schools refers to the total schools observed over the panel, then if a pupil from the reference cohort (4th grade 2005) is in a different school in 2004 or 2006, that school it would be counted here. **(iv)** School Socioeconomic Level variable: 0 Low; 1 Mid Low; 2 Middle; 3 Middle High; 4 High.

Table 4.8: Distribution of schools by number of classes per grade
Mini panel Cohort 1

No. classes per grade	3th Grade 2004		4th Grade 2005		5th Grade 2006	
	Fequency	Percent	Fequency	Percent	Fequency	Percent
1	6,047	73%	6,103	73%	5,906	73%
2	1,529	18%	1,573	19%	1,587	20%
3	505	6%	489	6%	458	6%
4	143	2%	125	1%	124	2%
5	28	0%	27	0%	27	0%
6 or more	21	0%	21	0%	20	0%
Total	**8,273**	**100%**	**8,338**	**100%**	**8,122**	**100%**
Sub Group	**2,177**	**26%**	**2,187**	**26%**	**2,169**	**27%**

Note: Some Schools change the number of schools between 3^{rd} and 5^{th} grade but it does not change dramatically total composition.

4.3.1 Sample selection criteria

The main objective of this chapter is to investigate non-random assignment of students to teachers, and as shown in Table 4.8, the group of schools which potentially could apply grouping strategies is approximately 27%.

Next, we present the selection criteria considered to create the sample in which we test for evidence of pupil sorting.

1^{st} Sample selection criterion: School selection

We select an initial sub group of schools with two to three classes per grade. There are not many schools with more than three classes per grade (around 2% of schools) and as we increase the number of classes per grade the testing methodologies get even more complex.

Simultaneously and for simplicity of the sorting analysis, we require schools to have the same number of classes per cohort over the three year period. So, in Table 9 we present the distribution of the schools which fulfil this first selection criterion. The selected represent approximately 21% of total schools.

Table 4.9: First criterion: School selection

	No. classes per grade	3th Grade 2004		4th Grade 2005		5th Grade 2006	
		Fequency	Percent	Fequency	Percent	Fequency	Percent
Cohort 1	2	1,319	77%	1,319	77%	1,319	77%
	3	384	23%	384	23%	384	23%
	Total	**1,703**	**100%**	**1,703**	**100%**	**1,703**	**100%**
	% of Total	**21%**		**20%**		**21%**	

Note: The 21% and 20% are with respect to the total number of schools shown in Table 4.8.

Given the methodology proposed for testing non-random assignment, we need schools with the same number of classes every year in the cohort under analysis. Thus, we separate the first school selection into two groups: those which have two classes per grade (**Group 1**) and those with three classes per grade (**Group 2**).

We check the distribution of class letter, or class identifier within school-grade, for both groups.[3] In Table 4.10, we see there are few cases which their class letter do not correspond neither "A" nor "B" as is expected for schools with two classes per grade. In Table 4.11 we also identify few observations of classes with unexpected class letters, different from "A", "B" and "C".

Table 4.10: Distribution of students by class letter
Group 1 (2 classes per grade)

		3th Grade 2004		4th Grade 2005		5th Grade 2006	
	Class Letter	Fequency	Percent	Fequency	Percent	Fequency	Percent
Cohort 1	A	42,722	51%	44,590	50%	42,911	51%
	B	41,602	49%	43,591	49%	41,791	49%
	C	59	0%	63	0%	27	0%
	D	50	0%	51	0%	52	0%
	E	29	0%	29	0%	27	0%
	F & others	26	0%	69	0%	31	0%
	Total	**84,488**	**100%**	**88,393**	**100%**	**84,839**	**100%**

Note: Classes classified with letters different from A or B are going to be re-labelled. We try to keep track of those classes who were classified with a particular letter in 3rd grade maintain the same classification for the rest of the period. In schools with just two classes per grade it does not make sense having a class with letter C or higher, we assume it could be a typo.

Table 4.11: Distribution of students by class letter
Group 2 (3 classes per grade)

		3th Grade 2004		4th Grade 2005		5th Grade 2006	
	Class Letter	Fequency	Percent	Fequency	Percent	Fequency	Percent
Cohort 1	A	13,111	34%	13,569	34%	13,173	34%
	B	12,986	33%	13,449	33%	12,929	33%
	C	12,756	33%	13,387	33%	12,787	33%
	D	25	0%	26	0%	27	0%
	Total	**38,878**	**100%**	**40,431**	**100%**	**38,916**	**100%**

Note: There is only one class in the sample classified with letter D. In every year, the other two class letters were identified as A and B, so we replace the letter D by C.

In Group 1, with two classes per grade, we still observe students in classes labelled as "C" or higher. Despite the fact they represent less than 1% of students, it is preferable to reallocate them into "A" and "B" classification. To do the reallocation we check case by case for appropriateness of re-labelling. The same methodology was followed for Group 2 which only had one unexpected class letter ("D").

2^{nd} Sample selection criterion: Student selection

Previously, when we selected cohorts and mini panels, we dropped students with observations out of the regular path from 3^{rd} to 5^{th} grade. Using as a reference cohort the students enrolled in the selected 4^{th} grade cohort, and we recover those

[3]Classes within school and grade are identified by letters (A, B, C, etc), depending on the number of classes per grade. Generally if there is only one class per grade, it is labeled as "A" (e.g 4^{th} grade A).

pupils who were observed in 3^{rd} grade the year before and in 5^{th} grade the year after.

After the initial student selection, we are sure all the observations correspond to pupils in the expected track from 3^{rd} to 5^{th}, but also including pupils from the reference cohort who have missing observation or who have repeated grades. Now, the concern is about having a sample where we observe pupils through the whole period in the mini panel.

We are forced to have a balanced mini panel, to reduce bias when testing sorting evidence. In Table 4.12, we present the distribution of students by the number of registers observed during the three years period.

After applying this selection criterion we continue the analysis with 80% of the original selected cohort in both groups. The dropped cases correspond to students who have repeated a grade or who have had one observation out of the expected path.

We also need to eliminate from the mini panel, the students who changed school during this period as they could be another source of bias when we test grouping evidence.

In Table 4.13 we present the pupils mobility distribution. We find a low change rate in both groups of schools, with around 3.5% in Group 1 and 1.2% in Group 2.

Table 4.12: No. registers per student within the mini panel

	No. registers per student	Group 1			Group 2		
		3rd 2004	4th 2005	5th 2006	3rd 2004	4th 2005	5th 2006
Cohort 1	1	5,411	2,312	5,715	2,704	1,054	3,013
	2	8,173	15,177	8,220	3,886	7,089	3,615
	3	70,904	70,904	70,904	32,288	32,288	32,288
	Total	**84,488**	**88,393**	**84,839**	**38,878**	**40,431**	**38,916**
	Sub group	*83.9%*	*80.2%*	*83.6%*	*83.0%*	*79.9%*	*83.0%*

Note: **(i)** Students with 3 registers through the period are those who follow the regular path from 3rd to 5th grade (no missing observations and no repeating grades). **(ii)** **Group 1:** Schools with 2 classes per grade from 3rd to 5th grade; **Group 2:** Schools with 3 classes per grade from 3^{rd} to 5^{th} grade.

Table 4.13: Student mobility in the mini panel

	School change	Group 1			Group 2		
		3rd 2004	4th 2005	5th 2006	3rd 2004	4th 2005	5th 2006
Cohort 1	No	68,414	68,425	70,904	31,911	31,919	32,288
	Yes	2,490	2,479	-	377	369	-
	Total	**70,904**	**70,904**	**70,904**	**32,288**	**32,288**	**32,288**
	Mobility Rate	*3.5%*	*3.5%*	*0.0%*	*1.2%*	*1.1%*	*0.0%*

Note: **(i)** We look for students who have not changed schools along the 3 year period **(ii)** **Group 1:** Schools with 2 classes per grade from 3rd to 5th grade; **Group 2:** Schools with 3 classes per grade from 3^{rd} to 5^{th} grade.

Considering only stayer pupils (or *never school movers*), we are able to track students within schools and check whether there exists evidence of grouping based on previous school marks.

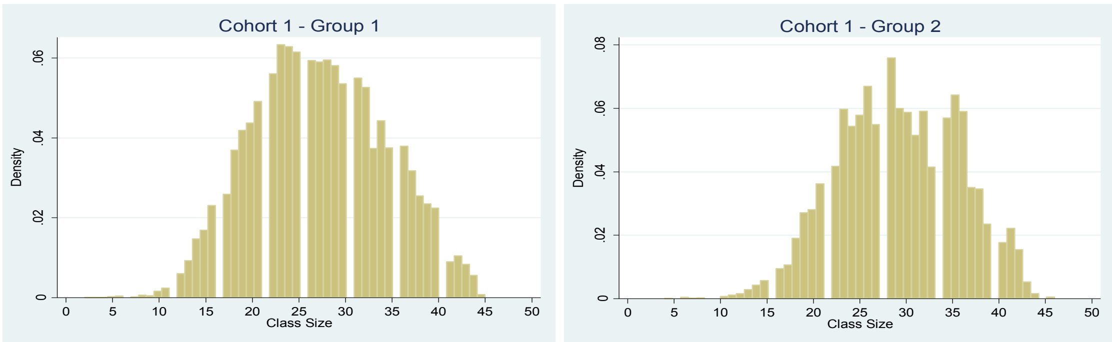
3^{rd} **Sample selection criterion: Class size selection**

The methodologies for testing non-random assignment consist basically in class distribution comparisons. From the real classes we create counterfactual classes to compare with.

The consistency of the estimators will depend on the number of observations we have per class. In the literature, the minimum number of pupils per class considered significant for different type of estimations varies between 10 and 20.

In Figure 4.1, we show the histograms of the class size for Cohort 1 - **Group 1** and **Group 2** after all previous selections. The number of students enrolled in classes with fewer than 15 pupils accumulate 3% of the remaining panel in **Group 1** and less than 0.5% in **Group 2**.

Figure 4.1: Class Size Histogram in selected groups



Note: **(i)** Both histograms correspond to the class size of the total classes observed along the panel (in 3^{rd}, 4^{th} and 5^{th} grade) for each group. **(ii) Group 1:** Schools with 2 classes per grade from 3^{rd} to 5^{th} grade; **Group 2:** Schools with 3 classes per grade from 3^{rd} to 5^{th} grade.

Holding the previous selection criteria, we keep schools with a minimum of 15 pupils per class, in all classes for the whole period. After applying the last selection criterion, we got a sample of 1,153 schools with 60,975 pupils every year.

4.3.2 Description of the mini panel samples

The mini panel samples are composed of students observed for three consecutive years in the same school, without repeating grades. The balanced panels are formed only for schools with two or three classes per grades along the whole period, represented by **Group 1** and **Group 2**, respectively.

Comparing the two sample groups in Table 4.14 with the original mini panel Cohort 1, described in Table 4.7, we do not observe big differences in the mean of average school marks, only the GPA increases by two decimal points approximately in both groups.

The average age of pupils does not change either, while the attendance slightly increases, and the proportion of pupils with special needs stays relatively low.

At class level, when we compare the availability of Language and Maths teacher register we find similar rates in our samples (around 0.98) than the observed in the original cohort. The number of schools where teachers are observed also stay very similar. However, average number of grades and classes decreased in both selected samples with respect to the original cohort. The proportion of SS teacher also increases from 0.30 to 0.38 and 0.39 in **Group 1** and **Group 2**, respectively.

Table 4.14: Summary statistics mini panel Cohort 1 - Groups 1 & 2

	Group 1				Group 2			
Variable	**Mean**	**Std.Dev.**	**Min**	**Max**	**Mean**	**Std.Dev.**	**Min**	**Max**
Pupil Level								
GPA	5.86	0.57	2.3	7	5.84	0.56	3.2	7
Average school Language marks	5.55	0.70	2.2	7	5.52	0.70	3.2	7
Average school Maths marks	5.44	0.78	2.2	7	5.41	0.77	2.8	7
Gender (Female=1)	0.51	0.50	0	1	0.50	0.50	0	1
Age	9.68	0.59	7	14.5	9.70	0.57	7	14.5
Attendance	94.11	6.55	0	100	94.21	5.23	44.3	100
Special Needs	0.01	0.10	0	1	0.02	0.10	0	1
Class Level								
Available Language teacher	0.99	0.12	0	1	0.98	0.14	0	1
Available Maths teacher	0.99	0.11	0	1	0.99	0.11	0	1
No. Schools per Language teacher	1.06	0.24	1	2	1.06	0.23	1	3
No. Schools per Maths teacher	1.07	0.25	1	3	1.05	0.23	1	3
No. Grades per Language teacher	1.60	0.93	1	7	1.52	0.83	1	7
No. Grades per Maths teacher	1.67	1.00	1	7	1.54	0.88	1	6
No. Classes per Language teacher	1.60	0.93	1	7	1.52	0.83	1	7
No. Classes per Maths teacher	2.26	1.77	1	11	2.34	1.86	1	10
Subject Specialist teacher	0.38	0.49	0	1	0.39	0.49	0	1
School Level								
Municipal schools	0.47	0.50	0	1	0.58	0.49	0	1
Private Voucher schools	0.47	0.50	0	1	0.30	0.46	0	1
Unsubsidised Private schools	0.06	0.24	0	1	0.12	0.33	0	1
Rural Area	0.05	0.21	0	1	0.01	0.12	0	1
School Socioeconomic Level	2.05	0.92	0	4	2.17	0.93	0.3	4
Number of students per grade	67.63	12.63	37.3	93	104.22	17.46	60	137
Number of students per class	26.67	6.07	15	45	28.38	5.49	16.3	41
Number of students	60,975				29,881			
Number of classes	6,918				3,195			
Number of schools	1,153				355			

Notes: **(i) Group 1:** Schools with 2 classes per grade from 3^{rd} to 5^{th}; Group 2: Schools with 3 classes per grade from 3^{rd} to 5^{th} grade. **(ii)** The number of students corresponds to the total pupil observed in Groups 1 and 2. **(iii)** The number of classes corresponds to all the classrooms observed over the three year periods in all schools. **(iv)** The number of schools refers to the total schools observed over the panel, then if a pupil from the reference cohort (4^{th} grade 2005) is in a different school in 2004 or 2006, that school it would be counted here. **(v)** School Socioeconomic Level variable: 0 Low; 1 Mid Low; 2 Middle; 3 Middle High; 4 High.

In terms of school distributions, the participation of *Municipal* schools is reduced from 0.60 to 0.47 in **Group 1** and 0.58 in **Group 2**. *Private Voucher* schools increase their participation only in **Group 1** from 0.35 to 0.47, while in **Group 2** decreases to 0.30. The proportion of *Unsubsidised Private* schools stay the same in **Group 1**, but it is the double (0.12) in **Group 2**. The distributions of students by type of school dependence are presented in Table 4.15.

Table 4.15: Distribution of students by type of school dependence in mini panel Cohort 1 - Groups 1 & 2

		Group 1			Group 2		
		3rd 2004	4th 2005	5th 2006	3rd 2004	4th 2005	5th 2006
Cohort 1	Municp	26,881	26,852	26,852	16,867	16,867	16,867
		44%	44%	44%	56%	56%	56%
	Priv. Vouch.	30,633	30,987	31,024	9,772	9,917	9,917
		50%	51%	51%	33%	33%	33%
	Unsubs. Priv.	3,242	3,097	3,097	3,242	3,097	3,097
		6%	5%	5%	11%	10%	10%
	Total	**60,756**	**60,936**	**60,973**	**29,881**	**29,881**	**29,881**
		100%	**100%**	**100%**	**100%**	**100%**	**100%**

Note: **(i)** The table shows the number of students observed in each cohort by grade-year, and distributed by type of school dependence. **(ii)** The school dependence is classify as: Municp: *Municipal* schools; Priv. Vouch: *Private Voucher* schools; Unsubs. Priv: *Unsubsidised Private* schools.

In **Group 1**, *Private Voucher* schools represent the highest participation of pupils (around 51%), while in **Group 2** the *Municipal* schools concentrate the highest proportion of students enrolled with approximately 56% of the sample. Similar to the distribution in terms of schools, *Unsubsidised Private* schools almost double their participation of students from 5% to 10%, when comparing **Group 1** and **Group 2**. **Group 1** then maintains similar characteristics with respect to the original 4^{th} cohorts than **Group 2**.

However, it is important to note that the size of **Group 2** is almost a third of **Group 1** in terms of schools, but it is just below the half in terms of pupils.

4.4 Evidence regarding non-random assignment

The methodology used to analyse potential student sorting considers the construction of counterfactual classes within schools, and compares these to the real classes. The counterfactuals are organised in two categories:

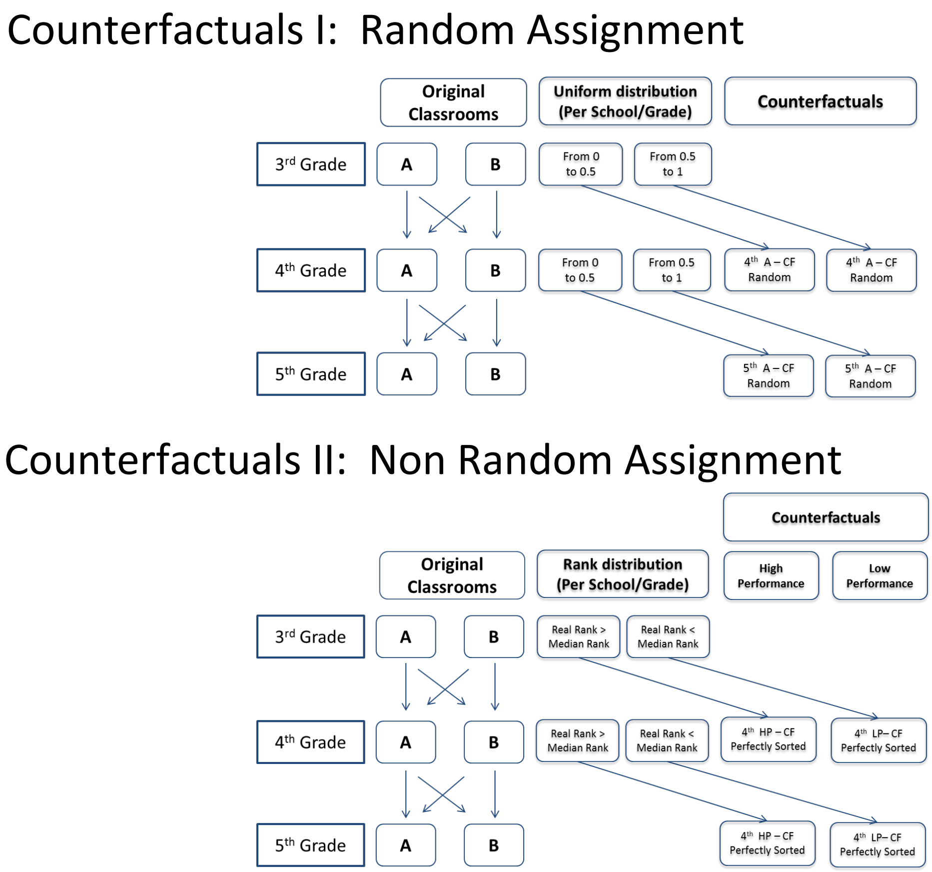
1. **Random assigned counterfactuals (RDM)**
2. **Perfectly sorted counterfactuals (SRT)**

The first category represents random assignment process, and is created by assigning students randomly into a uniform distribution for grades 3^{rd} and 4^{th}. From the randomly assigned distribution of pupils per cohort-grade, we generate two additional random classes (RDM) per grade equally distributed. The second category refers to non-random assignments of pupil to classrooms. These counterfactual classes are created discretionally ranking students within school based on previous year school marks (e.g. GPA, Language, Maths), which are observed by principals. Once pupils are artificially sorted we generate two type of classes; (i) the "High Performance" (HP), with the better ranked pupils, and (ii) the "Low Performance" (LP), with students from the bottom of the distribution. Hence, in every comparable grade (4^{th}, 5^{th}) we will have two set of artificially created counterfactuals.

1. **RDM-CF**: *"4A-RDM", "4B-RDM"; "5A-RDM", "5B-RDM"*
2. **SRT-CF**: *"4-HP-CF", "4-LP-CF"; "5-HP-CF", "5-LP-CF"*

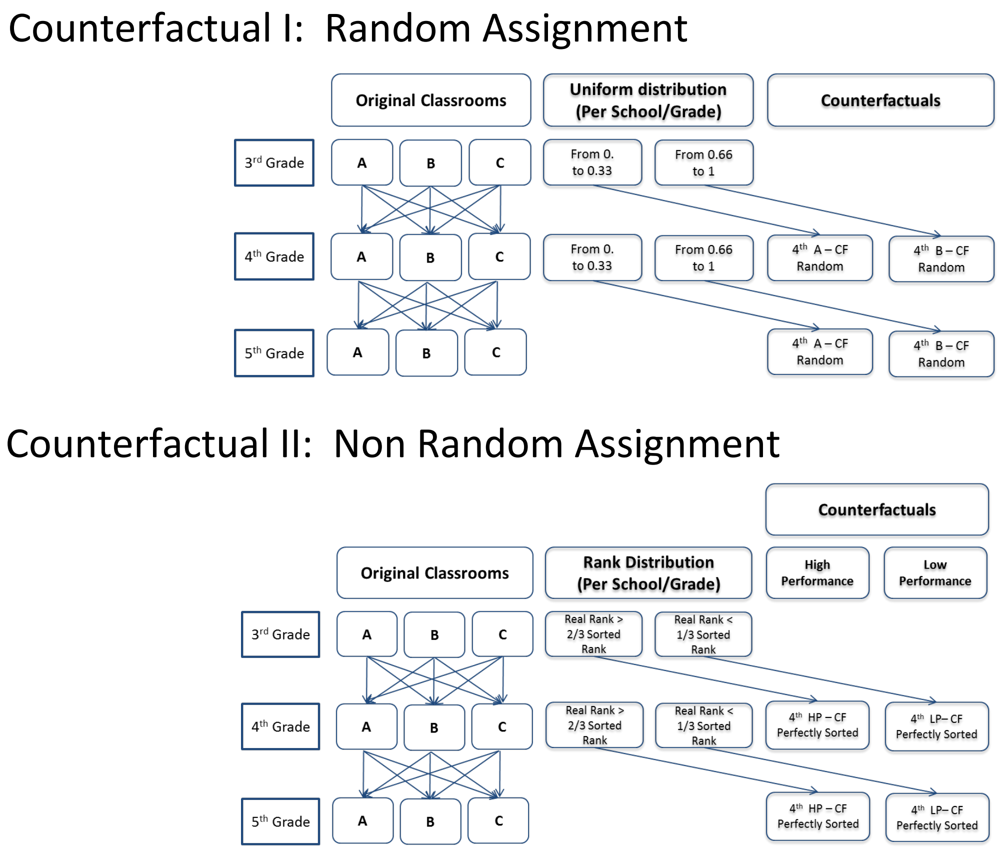
Figure 4.2 shows graphically the generation process of the two counterfactual categories when we carry out a semi-experimental analysis for schools with two classes per grade.

Figure 4.2: Counterfactual classes generating process - Group 1



Notes: **(i) Counterfactual I:** Pupils are uniformly distributed into classes. **(ii) Counterfactual II:** Pupils are perfectly sorted distributed into classes based on previous school marks.

Figure 4.3: Counterfactual classes generating process - Group 2



Notes: **(i) Counterfactual I:** Pupils are uniformly distributed into classes. **(ii) Counterfactual II:** Pupils are perfectly sorted distributed into classes based on previous school marks.

In Figure 4.3 the generation of counterfactual groups slightly change as we are considering schools with three classes per grade. We still create two counterfactual classes per random assignment and another two for the perfectly sorted assignment. However, in both cases we only consider a third of the total distribution of pupils, either for the randomly assigned classes or the perfectly sorted classes. The perfectly sorted counterfactuals are generated with the lowest and the highest third parts of the sorted previously school marks distribution.

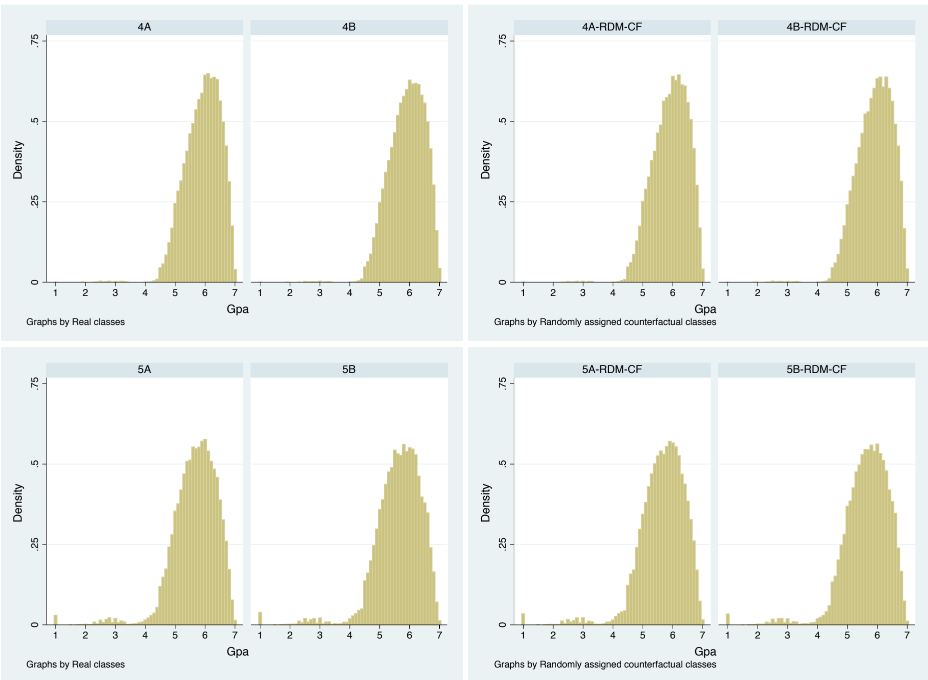
In the following subsections we compare the distributions of real and counterfactual classes graphically, and we test statistically whether it is possible to reject the hypothesis of non-random assignment within the Chilean education system given the evidence found in these two groups of schools.

4.4.1 Graphical analysis of Non-random assignment

We separate the graphical analysis in two approaches. The first approach shows *histogram* graphs comparing distributions based on GPA in real classes versus both types of counterfactuals: random (RDM) and perfectly sorted (SRT) counterfactuals. The second approach uses *kernel distribution* graphs comparing distributions of Language and Maths school marks in real classes versus the RDM and SRT counterfactual classes. The objective under the two approaches is to determine any visible evidence of student grouping in real classes distributions.

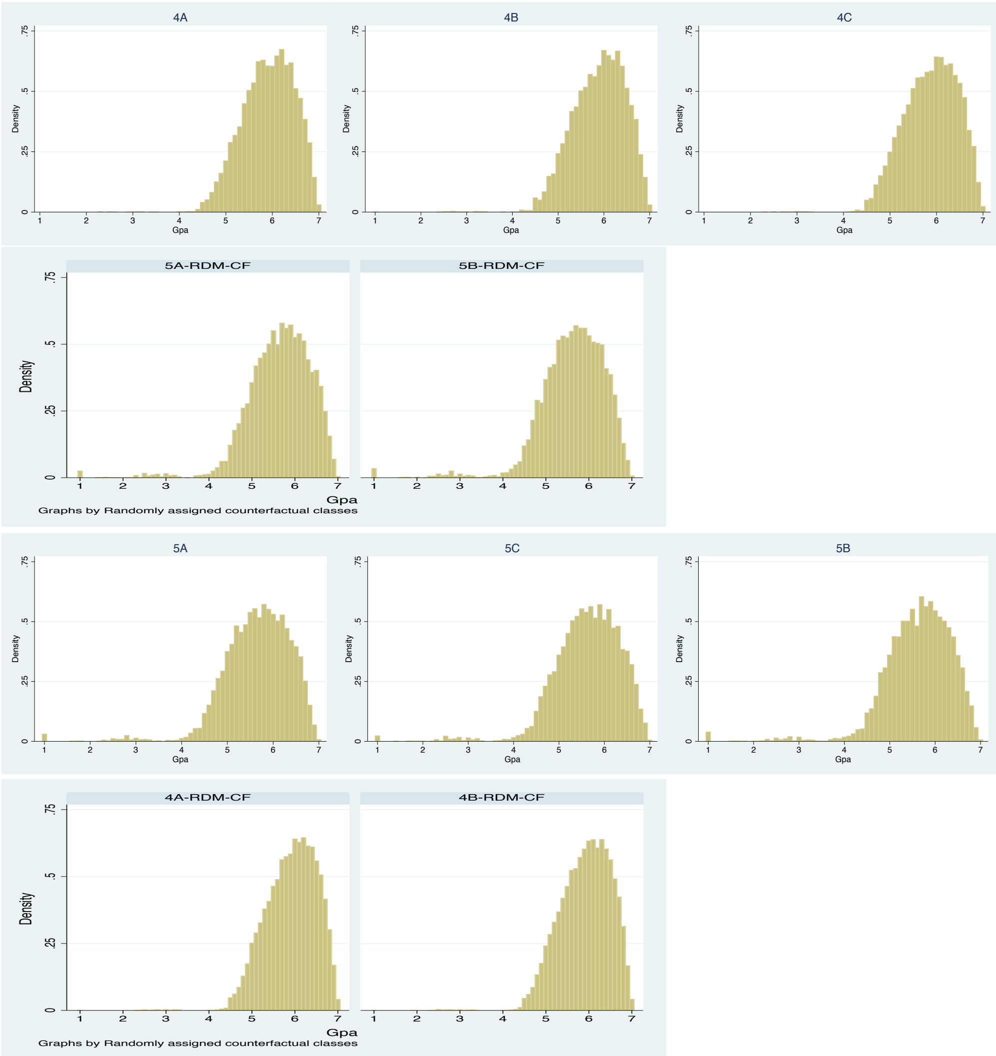
Considering the first graphical approach (*histogram*) and the comparison of real classes with RDM counterfactuals, we observe in Figure 4.4 how graphs of real classes in **Group 1** ('4A","4B";"5A","5B") do not differ from the graphs corresponding to randomly formed counterfactuals ('4A-RDM","4B-RDM";"5A-RDM","5B-RDM"). Similarly, for **Group 2**, in Figure 4.5 we compare the three real classes ('4A","4B", "4-C";"5A","5B", "5-C") with the same randomly created counterfactuals ('4A-RDM","4B-RDM";"5A-RDM","5B-RDM"). None of the class distributions (real and random counterfactuals) seems to differ among them, suggesting that pupils are randomly sorted into classes, at least from the school marks perspective.

Figure 4.4: Real classes vs random counterfactuals based on GPA
Group 1



Note: RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes.

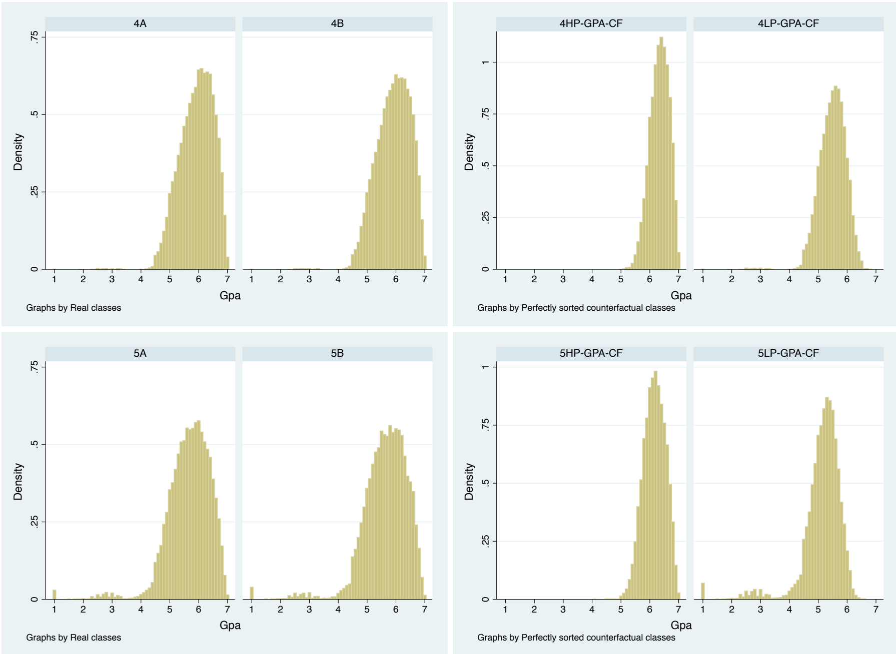
Figure 4.5: Real classes vs random counterfactuals based on GPA
Group 2



Note: RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes.

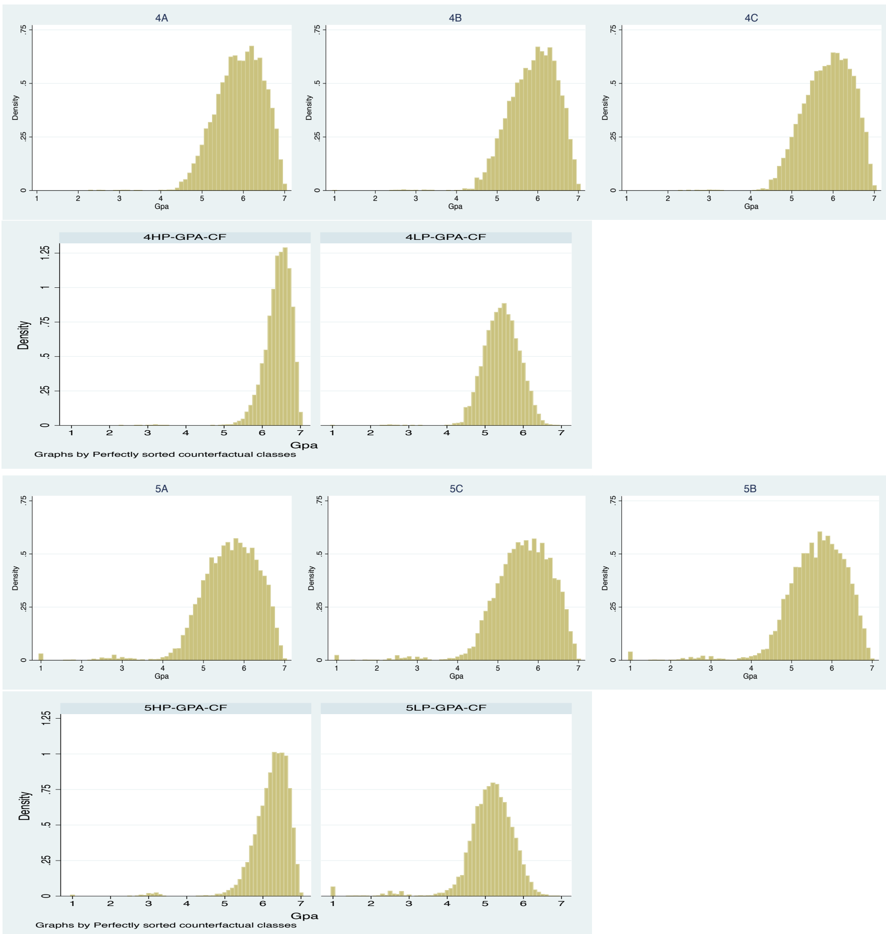
As an alternative, we also compare real classes to High and Low Performance counterfactuals (HP,LP). Given the graphs shown above, we expect to observe significant differences between real classes and perfectly sorted counterfactuals distributions based on GPA. Neither from **Group 1**, in Figure 4.6, or from **Group 2**, in Figure 4.7, we confirm that real classes do not appear to have similar histogram shapes in comparison to HP and LP counterfactuals. This finding supports the hypothesis of random assignment of students to classrooms as the real classes seem to significantly differ from perfectly sorted counterfactuals.

Figure 4.6: Real classes vs perfectly sorted counterfactuals based on GPA
Group 1



Notes: There are two perfectly sorted counterfactual classes per grade: **(i) HP-GPA-CF:** High Performance classes based on GPA. **(ii) LP-GPA-CF:** Low Performance classes based on GPA

Figure 4.7: Real classes vs perfectly sorted counterfactuals based on GPA
Group 2



Notes: There are two perfectly sorted counterfactual classes per grade: **(i) HP-GPA-CF:** High Performance classes based on GPA. **(ii) LP-GPA-CF:** Low Performance classes based on GPA

The *histogram* approach do not suggest grouping evidence based on previous year GPA when we compare the two groups of schools with the two artificially created counterfactuals. Our second graphical approach focuses on *kernel distributions*, where we compare real classes in **Group 1** and **Group 2** with both counterfactuals (RDM, SRT). Instead of showing comparisons based previous GPA, we compare real classes distributions with perfectly sorted counterfactuals based on previous Language and Maths school marks.[4]

Language and Maths kernel distributions

We compare school marks *kernel distributions* of real classes with school from the two type of counterfactuals (RDM, SRT). The perfectly sorted counterfactuals classes (HP, LP) are based either on previous Language or Maths school marks. To find supportive evidence of random assignment of pupil to classes we need that real *kernel distributions* are similar to their RDM counterfactuals, and significantly different from the SRT counterfactuals.

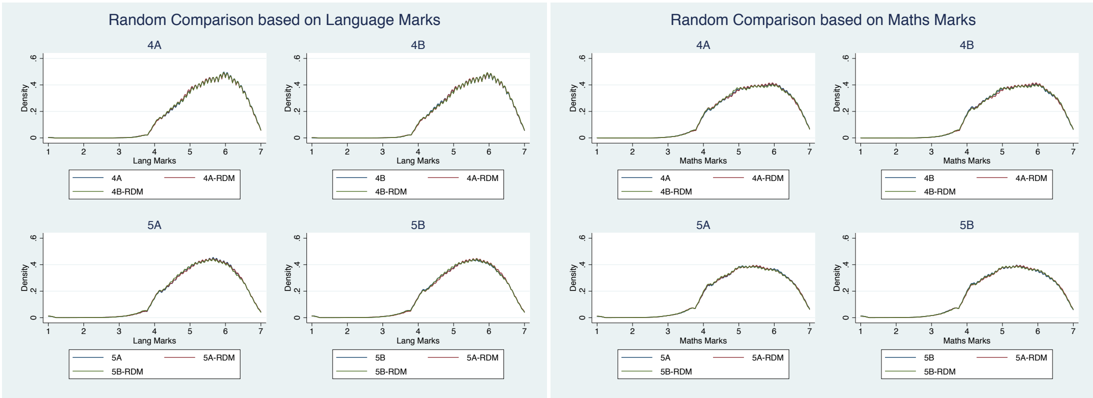
From **Group 1**, in Figure 4.8, we present comparisons of all real classes distributions, and their respective random counterfactuals (A-RDM, B-RDM) by subject. The do not seem to differ in any case. In Figure 4.9, we analyse the case for **Group 2**, where we do not observe differences between real classes and random counterfactuals either. These results are similar to what we found using the histogram graphs. Hence, there is no clear graphical evidence of non-random assignment of students to classes.

When we compare real classes with perfectly sorted random counterfactuals we expect to observe noticeable differences between *kernel distributions*, otherwise it might be some graphical evidence of non-random of pupil to school based on previous school marks. In Figure 4.10, we show the comparison of real classes distributions in **Group 1** for both subjects. In all four cases (4A, 4B, 5A, 5B) *kernel distributions* of SRT counterfactuals (HP-SRT, LP-SRT) are significantly different from real classes.

Similarly, when we make this comparison for **Group 2**, we see Figure 4.11 there are no similarities between real classes (4A, 4B, 4C, 5A, 5B, 5C) and the perfectly sorted random counterfactuals, for both Language and Maths subjects.

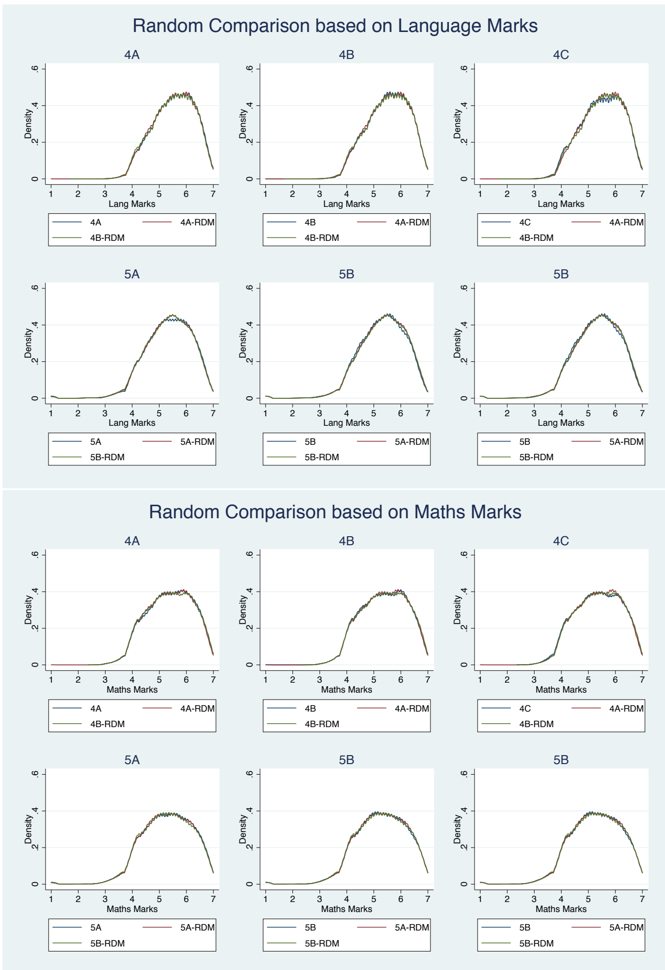
[4]Although we have made the comparisons of *kernel distributions* based on previous GPA, we do not report them as their conclusions are the same to those obtained from *histogram* graphical approach and the *kernel distribution* comparisons by subject. There are no differences between real classes and random counterfactual classes in GPA distributions, but there are significant differences between real classes and High Performance (HP) and Low Performance (LP) counterfactuals.

Figure 4.8: Real classes vs random counterfactuals based previous school marks Group 1



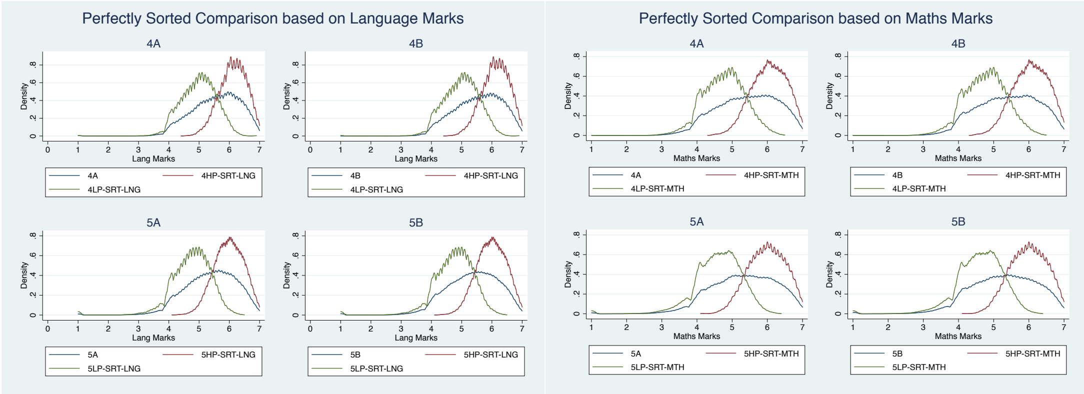
Notes: (i) RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes. **(ii)** In each plot we compare real classes **(4A, 4B, 5A, 5B)** with their respective random counterfactuals **(A, B RDM-CF)**

Figure 4.9: Real classes vs random counterfactuals based previous school marks Group 2



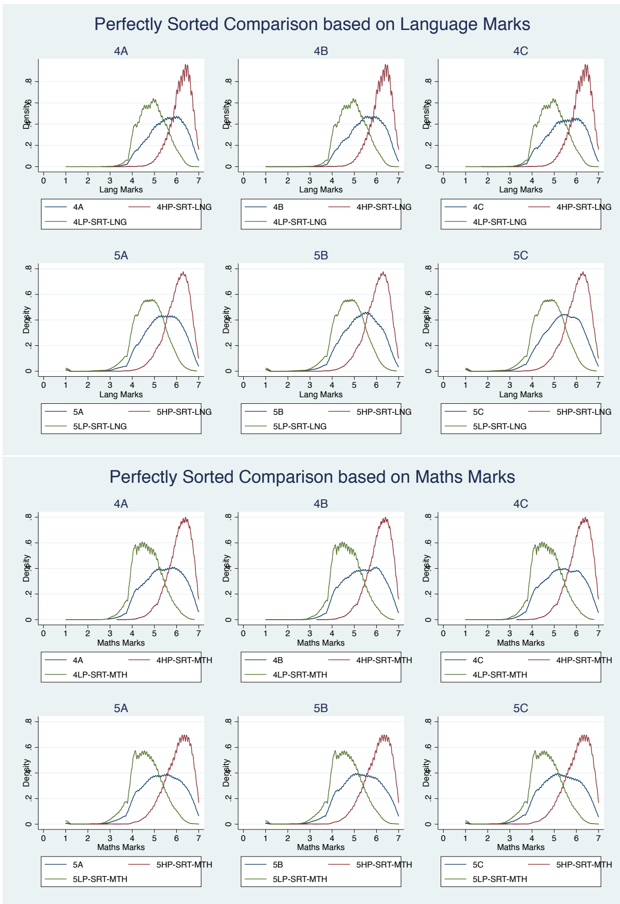
Notes: (i) RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes. **(ii)** In each plot we compare real classes **(4A, 4B, 4C, 5A, 5B, 5C)** with their respective random counterfactuals **(A, B RDM-CF**)

Figure 4.10: Real classes vs perfectly sorted counterfactuals based previous school marks
Group 1



Notes: There are two perfectly sorted counterfactual classes per and grade. **(i)** Based on Language Marks; **HP-SRT-LNG:** High Performance classes; **LP-SRT-LNG:** Low Performance classes. **(ii)** Based on Maths Marks; **HP-SRT-MTH:** High Performance classes; **LP-SRT-MTH:** Low Performance classes. **(iii)** In each plot we compare real classes **(4A, 4B, 5A, 5B)** with their respective random counterfactuals (**HP-SRT-LNG, HP-SRT-MTH; LP-SRT-LNG, LP-SRT-MTH)**

Figure 4.11: Real classes vs perfectly sorted counterfactuals based previous school marks
Group 2



Notes: There are two perfectly sorted counterfactual classes per and grade. **(i)** Based on Language Marks; **HP-SRT-LNG:** High Performance classes; **LP-SRT-LNG:** Low Performance classes. **(ii)** Based on Maths Marks; **HP-SRT-MTH:** High Performance classes; **LP-SRT-MTH:** Low Performance classes. **(iii)** In each plot we compare real classes **(4A, 4B, 4C, 5A, 5B, 5C)** with their respective random counterfactuals **(HP-SRT-LNG, HP-SRT-MTH; LP-SRT-LNG, LP-SRT-MTH)**

Although, graphical analysis is a good starting point, it is still necessary to go further and perform strict statistical tests to support our hypothesis of random assignment of student to classroom within schools.

4.4.2 Statistical testing of Non-random assignment

Even though graphical evidence supports the random assignment assumption of pupil to classrooms in the Chilean school system, we test statistically how consistent this hypothesis is for our two selected groups of schools. Similar to the graphical approachers, we make comparisons between real classes and the two type of counterfactuals based on the previous year school marks. If there is non-random assignment of student to classes, there should not be statistical differences between real classes and random (RDM) counterfactual classes, while we expect to find statistical difference between real classes and perfectly sorted (SRT) counterfactuals.

The comparisons between real classes and their counterfactuals are carried out with two statistical tests: (i) **T-test**, for measuring mean difference between real classrooms and counterfactuals; and (ii) **Kolgomorov-Smirnov (KS) test**, for assessing statistical difference between the distributions.[5] We test for "Sorting Evidence" (SoE) within school based on each possible comparison, given: the type of counterfactual, the type of statistical test, and the previous school performance measures which are used to create the RDM and SRT counterfactuals. In total, we have 12 independent non-random assignment measures per schools, represented by the *Non-Random Assignment* (*Non-RA*) Indexes, from (1) to (12), as it is described in Table 4.16.

Depending on the SoE observed in each school for every type of comparison, the *Non-RA*) Index classifies schools into five levels of non-random assignment: *None, Low, Medium, Med-High,* and *High*. Therefore, we sum the number of schools observed in each category, and analyse their distribution for both groups of school.

We compute the SoE to construct the *Non-RA*) Indexes for all schools in **Group 1** and **Group 2**. Table 4.17 shows how the SoE is estimated for the *Non-RA*) Index (1) to (6) in **Group 1**. Here, the type of counterfactual class used the RDM, which is tested either under **t-test** or **ks-test**. For every Index, the maximum possible SoE is **8**, meaning there is evidence of non-random assignment in the eight possible comparisons for this group of schools. Similarly, for **Group 2**, we present in Table 4.18 the maximum SoE that can be reached in this group

[5]The Kolmorov-Smirnov test is considered as the most appropriate test for comparing distributions (Gibbons and Chakraborti (2011)). The criterion comparison is stricter than a t-test as we are now comparing the whole distribution instead of just the mean.

when we compare real classes with RDM counterfactuals. Here, we have four more possible comparisons, therefore the maximum SoE score is **12**.

Table 4.16: Non-random assignment (Non-RA) Indexes per school

Type of Counterfactual Class	Type of Statistic Test	Performance measures	Non-RA Indexes
Random (RDM)	T-test	GPA	*Non-RA Index (1)*
		Language	*Non-RA Index (2)*
		Maths	*Non-RA Index (3)*
	KS-test	GPA	*Non-RA Index (4)*
		Language	*Non-RA Index (5)*
		Maths	*Non-RA Index (6)*
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*
		Language	*Non-RA Index (8)*
		Maths	*Non-RA Index (9)*
	KS-test	GPA	*Non-RA Index (10)*
		Language	*Non-RA Index (11)*
		Maths	*Non-RA Index (12)*

Note: **(i)** In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools (Group 1, Group 2) **(ii)** There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test.

Table 4.17: Potential cases of non-random assignment tested with RDM counterfactuals
Group 1

Real Class	Counterfactual Class	Hypothesis test	Sorting Evidence (SoE)	Condition
4A	4A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	4B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
4B	4A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	4B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
5A	5A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	5B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
5B	5A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	5B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
Maximum SoE Score *Non-RA Indexes (1) - (6)*			**8**	

Notes: **(i)** For schools in Group 1, the potential maximum evidence of non-random assignment is **8**. **(ii)** Every real class within school is compared with the two random counterfactual classes per grade **(A,B RDM)**. **(iii)** We use two statistical tests: T-test to compare means, and KS-test to compare distributions. **(iv)** We claim there is **sorting evidence (SoE)** in a particular comparison when the Null Hypothesis (Ho) of No differences between the classes is rejected at a 5% significance level (for both t-test and ks-test). **(v)** In the Hypothesis test column the Ho in brackets refers to the ks-test.

Table 4.18: Potential cases of non-random assignment tested with RDM counterfactuals
Group 2

Real Class	Counterfactual Class	Hypothesis test	Sorting Evidence (SoE)	Condition
4A	4A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	4B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
4B	4A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	4B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
4C	4A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	4B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
5A	5A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	5B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
5B	5A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	5B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
5C	5A-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
	5B-RDM	Ho: Mean difference = 0	**1**	**If t-tests (ks-test) rejects Ho**
		(Ho: Diff. in distribution = 0)	0	If t-tests (ks-test) does not reject Ho
Maximum SoE Score *Non-RA Indexes (1) - (6)*			**12**	

Notes: (i) For schools in Group 1, the potential maximum evidence of non-random assignment is **12**. **(ii)** Every real class within school is compared with the two random counterfactual classes per grade **(A,B RDM)**. **(iii)** We use two statistical tests: T-test to compare means, and KS-test to compare distributions. **(iv)** We claim there is **sorting evidence (SoE)** in a particular comparison when the Null Hypothesis (Ho) of No differences between the classes is rejected at a 5% significance level (for both t-test and ks-test). **(v)** In the Hypothesis test column the Ho in brackets refers to the ks-test.

The hypothesis testing suggests there is SoE when the hypothesis null (Ho) of mean difference between real and counterfactual classes equals to zero is rejected at a 5% significance level, when we apply the **t-test**. Similarly, it also suggests there is SoE when the Ho of difference in distributions between classes equals to zero is rejected at a 5% significance level, when using the **ks-test**.

The second round of comparisons, which corresponds to the construction of the *Non-RA*) Indexes (7) to (12), uses as counterfactual classes the SRT classes. The SRT counterfactuals are defined as High Performance (HP) and Low Performance (LP) based on the previous school marks. Similar to the comparisons shown above for the RDM counterfactuals, in this case we also separate the estimation of SoE and the construction *Non-RA*) Index by group of schools.

In Table 4.19, we present all possible cases to test for SoE in **Group 1** based on the comparisons between real classes and the SRT counterfactuals. The difference with respect to the RDM comparison (Table 4.17) is with respect to the hypothesis test, as in this case we account for SoE when we DO NOT reject the null hypothesis of mean differences equal to zero or distribution differences equal to zero, when we test with **t-test** and **ks-test**, respectively.

Table 4.19: Potential cases of non-random assignment tested with SRT counterfactuals
Group 1

Real Class	Counterfactual Class	Hypothesis test	Sorting Evidence (SoE)	Condition
4A	4A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	4B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
4B	4A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	4B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
5A	5A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	5B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
5B	5A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	5B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
Maximum SoE Score *Non-RA Indexes (7) - (12)*			**8**	

Notes: **(i)** For schools in Group 1, the potential maximum evidence of non-random assignment is **8**. **(ii)** Every real class within school is compared with the two perfectly sorted counterfactual classes per grade **(A,B SRT)**. **(iii)** We use two statistical tests: T-test to compare means, and KS-test to compare distributions. (iv) We claim there is **sorting evidence (SoE)** in a particular comparison when the Null Hypothesis (Ho) of No differences between the classes is NOT rejected at a 5% significance level (for both t-test and ks-test). **(v)** In the Hypothesis test column the Ho in brackets refers to the ks-test.

Table 4.20 shows all the combinations to compare real classes with SRT counterfactuals in **Group 2**. As we described earlier for the RDM comparison in Table 4.18, in this group of school the maximum SoE score which can be reached per schools is **12**. All testing made for the *Non-RA* Indexes (7) to (12) are looking for NON rejection of the Ho which states non mean difference when we apply the **t-test** and non distribution differences when we use the **ks-test**.

Table 4.20: Potential cases of non-random assignment tested with SRT counterfactuals
Group 2

Real Class	Counterfactual Class	Hypothesis test	Sorting Evidence (SoE)	Condition
4A	4A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	4B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
4B	4A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	4B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
4C	4A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	4B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
5A	5A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	5B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
5B	5A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	5B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
5C	5A-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
	5B-SRT	Ho: Mean difference = 0	0	If t-tests (ks-test) rejects Ho
		(Ho: Diff. in distribution = 0)	**1**	**If t-tests (ks-test) does not reject Ho**
Maximum SoE Score *Non-RA Indexes (7) - (12)*			**12**	

Notes: (i) For schools in Group 1, the potential maximum evidence of non-random assignment is **12**. **(ii)** Every real class within school is compared with the two perfectly sorted counterfactual classes per grade **(A,B SRT)**. **(iii)** We use two statistical tests: T-test to compare means, and KS-test to compare distributions. (iv) We claim there is **sorting evidence (SoE)** in a particular comparison when the Null Hypothesis (Ho) of No differences between the classes is NOT rejected at a 5% significance level (for both t-test and ks-test). **(v)** In the Hypothesis test column the Ho in brackets refers to the ks-test.

Aggregating all (*Non-RA*) Indexes at school group level, we are able to count how many schools are classified in each category in every group. Depending of the distribution of schools among the categories, we could suggest whether there is *Low, Medium* or *High* evidence of non-random assignment of pupil to classes within schools.

The *Non-RA* Indexes (from (1) to (12)) take the following correspondence functional form when the comparisons between real classes and counterfactual classes are run in a school s from **Group 1 (G1)**. Schools from this group can have a maximum score of **8** in each *Non-RA* Index, as it is total possible combinations to compare between real classes and the counterfactuals.

$$
Non-RA_{s(G1)} = \begin{cases} \textit{None} & \text{if} \quad SoE_{s(G1)} = 0 \\ \textit{Low} & \text{if} \quad SoE_{s(G1)} = \{1,2\} \\ \textit{Medium} & \text{if} \quad SoE_{s(G1)} = \{3,4\} \\ \textit{Med-High} & \text{if} \quad SoE_{s(G1)} = \{5,6\} \\ \textit{High} & \text{if} \quad SoE_{s(G1)} = \{7,8\} \end{cases} \tag{4.1}
$$

The following functional form corresponds when the comparisons are applied in a school s from **Group 2 (G2)**, where the maximum SoE score is **12** given the possible combinations to compare real and counterfactual classes.

$$
Non-RA_{s(G2)} = \begin{cases} \textit{None} & \text{if} \quad SoE_{s(G2)} = 0 \\ \textit{Low} & \text{if} \quad SoE_{s(G2)} = \{1,2,3\} \\ \textit{Medium} & \text{if} \quad SoE_{s(G2)} = \{4,5,6\} \\ \textit{Med-High} & \text{if} \quad SoE_{s(G2)} = \{7,8,9\} \\ \textit{High} & \text{if} \quad SoE_{s(G2)} = \{10,11,12\} \end{cases} \tag{4.2}
$$

Both groups of school require that none of the comparisons provides SoE in order to be classified as a school with *None* evidence of non-random assignment. If schools from **Group 1** present 1 or 2 cases of SoE, they are classify as schools with *Low* evidence of non-random assignment, while schools from **Group 2** require from 1 to 3 cases to be in the same group. The same logic is used to identify schools from *Medium* to *High* levels of SoE. Therefore, we produce 5 different categories of school regarding the non-random assignment evidence of pupil to classrooms.

In the following subsection, we present the results of the means comparison **t-test**, and the non-parametric estimate of the difference in distributions **ks-test**. We aggregate the results at the school group level.

4.4.3 Measuring Non-random assignment within Group 1

Considering the graphical analyses from previous subsections, we expect to find low levels of non-random assignments in **Group 1** when we compare real classes to our artificially created counterfactuals.

As we have mentioned above, we designed 12 *Non-RA* Index per schools, where we classify every school into a category level of non-random assignment given the SoE within schools. Table 4.21 describes the total number of schools in **Group 1** (1,153), were classified in each *Non-RA* Index, from *None* to *High* evidence of non-random assignment of pupil to classrooms.

The results presented in Table 4.22 indicate how schools are distributed between the SoE categories given each *Non-RA* Index. As we can see, the results slightly vary within the type of counterfactual and statistical test used for every Index. In average, we observe that 69% of schools from **Group 1**, do not present any evidence of non-random assignment. While only 5% of schools within this group were classified with Medium or higher level of non-random assignment evidence.

Table 4.21: Estimated levels of Non-Random assignment
Group 1

Non-RA Indexes				Sorting Evidence (SoE)					Total No. of schools
				None (0)	Low (1-2)	Med (3-4)	Med-High (5-6)	High (7-8)	
Random (RDM)	T-test	GPA	*Non-RA Index (1)*	609	473	70	1	0	1,153
		Language	*Non-RA Index (2)*	664	435	54	0	0	1,153
		Maths	*Non-RA Index (3)*	742	376	35	0	0	1,153
	KS-test	GPA	*Non-RA Index (4)*	1,099	44	8	2	0	1,153
		Language	*Non-RA Index (5)*	1,090	51	12	0	0	1,153
		Maths	*Non-RA Index (6)*	1,102	48	2	1	0	1,153
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*	609	473	70	1	0	1,153
		Language	*Non-RA Index (8)*	914	206	33	0	0	1,153
		Maths	*Non-RA Index (9)*	957	176	20	0	0	1,153
	KS-test	GPA	*Non-RA Index (10)*	613	407	133	0	0	1,153
		Language	*Non-RA Index (11)*	563	432	158	0	0	1,153
		Maths	*Non-RA Index (12)*	634	382	137	0	0	1,153
			Non-RA Average Index	**800**	**292**	**61**	**0**	**0**	

Notes: **(i)** In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools **(ii)** There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test. **(v)** The level of Non-Random assignment depends on the **Sorting Evidence (SoE)** in every school: None if **SoE**=0; Low if **SoE**=1,2; Medium if **SoE**=3,4; Med-High if **SoE**=5,6; and High if **SoE**=7,8.

Table 4.22: Estimated levels of Non-Random assignment in percentage
Group 1

Non-RA Indexes				Sorting Evidence (SoE)					Total schools (%)
				None (0)	Low (1-2)	Med (3-4)	Med-High (5-6)	High (7-8)	
Random (RDM)	T-test	GPA	*Non-RA Index (1)*	53%	41%	6%	0%	0%	100%
		Language	*Non-RA Index (2)*	58%	38%	5%	0%	0%	100%
		Maths	*Non-RA Index (3)*	64%	33%	3%	0%	0%	100%
	KS-test	GPA	*Non-RA Index (4)*	95%	4%	1%	0%	0%	100%
		Language	*Non-RA Index (5)*	95%	4%	1%	0%	0%	100%
		Maths	*Non-RA Index (6)*	96%	4%	0%	0%	0%	100%
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*	53%	41%	6%	0%	0%	100%
		Language	*Non-RA Index (8)*	79%	18%	3%	0%	0%	100%
		Maths	*Non-RA Index (9)*	83%	15%	2%	0%	0%	100%
	KS-test	GPA	*Non-RA Index (10)*	53%	35%	12%	0%	0%	100%
		Language	*Non-RA Index (11)*	49%	37%	14%	0%	0%	100%
		Maths	*Non-RA Index (12)*	55%	33%	12%	0%	0%	100%
			Non-RA Average Index	**69.4%**	**25.3%**	**5.3%**	**0%**	**0%**	

Notes: **(i)** In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools **(ii)** There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test. **(v)** The level of Non-Random assignment depends on the **Sorting Evidence (SoE)** in every school: None if **SoE**=0; Low if **SoE**=1,2; Medium if **SoE**=3,4; Med-High if **SoE**=5,6; and High if **SoE**=7,8.

4.4.4 Measuring Non-random assignment within Group 2

Intuitively, we expected higher evidence of non-random assignment of pupil to classrooms in schools from **Group 2** rather than **Group 1**, as the chances of grouping increase having one additional class per grade.

Tables 4.23 and 4.24 show how the evidence of non-random assignment increases form the *None* level to the *Low* level category in **Group 2**. The pattern is observed in all *Non-RA* Indexes, when we compare real classes with both type of counterfactuals and using both statistical tests.

Table 4.23: Estimated levels of Non-Random assignment
Group 2

Non-RA Indexes				Sorting Evidence (SoE)					Total No. of schools
				None (0)	Low (1-3)	Med (4-6)	Med-High (7-9)	High (10-12)	
Random (RDM)	T-test	GPA	*Non-RA Index (1)*	176	141	34	4	0	355
		Language	*Non-RA Index (2)*	145	164	41	5	0	355
		Maths	*Non-RA Index (3)*	168	163	20	4	0	355
	KS-test	GPA	*Non-RA Index (4)*	244	102	8	1	0	355
		Language	*Non-RA Index (5)*	229	114	11	1	0	355
		Maths	*Non-RA Index (6)*	264	86	4	1	0	355
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*	176	141	34	4	0	355
		Language	*Non-RA Index (8)*	145	164	41	5	0	355
		Maths	*Non-RA Index (9)*	168	163	20	4	0	355
	KS-test	GPA	*Non-RA Index (10)*	111	191	49	4	0	355
		Language	*Non-RA Index (11)*	65	188	91	11	0	355
		Maths	*Non-RA Index (12)*	73	190	80	12	0	355
			Non-RA Average Index	**164**	**151**	**36**	**5**	**0**	

Note: **(i)** In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools (ii) There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test. **(v)** The level of Non-Random assignment depends on the **Sorting Evidence (SoE)** in every school: None if **SoE**=0; Low if **SoE**=1,2,3; Medium if **SoE**=5,4,6; Med-High if **SoE**=7,8,9; and High if **SoE**=10,11,12.

Table 4.24: Estimated levels of Non-Random assignment in percentage
Group 2

Non-RA Indexes				Sorting Evidence (SoE)					Total schools (%)
				None (0)	Low (1-3)	Med (4-6)	Med-High (7-9)	High (10-12)	
Random (RDM)	T-test	GPA	*Non-RA Index (1)*	50%	40%	10%	1%	0%	100%
		Language	*Non-RA Index (2)*	41%	46%	12%	1%	0%	100%
		Maths	*Non-RA Index (3)*	47%	46%	6%	1%	0%	100%
	KS-test	GPA	*Non-RA Index (4)*	69%	29%	2%	0%	0%	100%
		Language	*Non-RA Index (5)*	65%	32%	3%	0%	0%	100%
		Maths	*Non-RA Index (6)*	74%	24%	1%	0%	0%	100%
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*	50%	40%	10%	1%	0%	100%
		Language	*Non-RA Index (8)*	41%	46%	12%	1%	0%	100%
		Maths	*Non-RA Index (9)*	47%	46%	6%	1%	0%	100%
	KS-test	GPA	*Non-RA Index (10)*	31%	54%	14%	1%	0%	100%
		Language	*Non-RA Index (11)*	18%	53%	26%	3%	0%	100%
		Maths	*Non-RA Index (12)*	21%	54%	23%	3%	0%	100%
			Non-RA Average Index	**46.1%**	**42.4%**	**10.2%**	**1.3%**	**0%**	

Note: **(i)** In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools (ii) There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test. **(v)** The level of Non-Random assignment depends on the **Sorting Evidence (SoE)** in every school: None if **SoE**=0; Low if **SoE**=1,2,3; Medium if **SoE**=5,4,6; Med-High if **SoE**=7,8,9; and High if **SoE**=10,11,12.

However, we do not find schools with high levels of SoE, and we only find 1% of the total sub-sample being classify in the Med-High level of the *Non-RA* Index. It is worth noting that the distribution of schools in the None and Low category of non-random assignment pass from 94.7% in **Group 1** to 88.5% in **Group 2**. This difference is due to an approximated increase of 5% in the proportion of schools identified with *Medium* level of SoE.

4.4.5 Aggregated levels of Non-random assignment

Aggregating the measures of *Non-RA* Indexes for **Groups 1** and **2**, we obtain the estimated average of SoE for our whole sample. In Table 4.25, we show that in average the 64% of schools DO NOT present any evidence of non-random assignment. In the total, we only find in average a 7% percent of schools with *Medium* or higher levels of SoE. Theses aggregated results provide evidence enough of Low levels of non-random assignment of pupil to classroom based on previous school marks.[6]

Table 4.25: Estimated levels of Non-Random assignment in percentage Groups 1 & 2

Non-RA Indexes				Sorting Evidence (SoE)					Total schools (%)
				None	Low	Med	Med-High	High	
Random (RDM)	T-test	GPA	*Non-RA Index (1)*	52%	41%	7%	0%	0%	100%
		Language	*Non-RA Index (2)*	54%	40%	6%	0%	0%	100%
		Maths	*Non-RA Index (3)*	60%	36%	4%	0%	0%	100%
	KS-test	GPA	*Non-RA Index (4)*	89%	10%	1%	0%	0%	100%
		Language	*Non-RA Index (5)*	87%	11%	2%	0%	0%	100%
		Maths	*Non-RA Index (6)*	91%	9%	0%	0%	0%	100%
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*	52%	41%	7%	0%	0%	100%
		Language	*Non-RA Index (8)*	70%	25%	5%	0%	0%	100%
		Maths	*Non-RA Index (9)*	75%	22%	3%	0%	0%	100%
	KS-test	GPA	*Non-RA Index (10)*	48%	40%	12%	0%	0%	100%
		Language	*Non-RA Index (11)*	42%	41%	17%	1%	0%	100%
		Maths	*Non-RA Index (12)*	47%	38%	14%	1%	0%	100%
			Non-RA Average Index	**63.9%**	**29.3%**	**6.4%**	**0.3%**	**0%**	

Notes: (i) In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools (ii) There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test. **(v)** The level of Non-Random assignment depends on the **Sorting Evidence (SoE)** in every school: None if **SoE**=0; Low if **SoE**=1,2 (1,2,3); Medium if **SoE**=3,4 (5,4,6); Med-High if **SoE**=5,6 (7,8,9); and High if **SoE**=7,8 (10,11,12); in brackets conditions for **SoE** categories in Group 2.

[6]In Appendix 4.3, we present the summary tables of the SoE found in the other two selected cohorts. The results are similar to those we have discussed in this section.

4.5 Conclusion

A main concern when estimating teacher and school effects is the possible non-random assignment of teachers to classrooms, as this can result in inconsistent estimators. Particularly, we checked for evidence of non-random assignment sorting of students to teachers or classrooms.

We selected three 4^{th} grade cohorts (2005, 2007, 2009) and we created mini panels from them considering their 3^{rd} and 5^{th} grade in their previous and subsequent year. From each mini panel we select a specific sample of schools where we could actually test the evidence of student sorting. In this chapter, we choose one representative cohort (4^{th} grade 2005 - Cohort 1) to show in detail the sample selection process, and discuss the results obtained from the non-random assignment analysis.

We selected 28% out of the 8,338 schools observed for Cohort 1, as we immediately ruled out grouping behaviour in schools with only one class per grade. Therefore, we focused just on schools with two and three classes per grade (**Groups 1** and **2** respectively), restricting the sample to 20% of the total schools. We also applied some student and class selections, dropping repeating students and eliminating schools with classrooms smaller than 15 pupils.

The final sample consisted of 1,508 schools in Cohort 1, 77% with two classes per grade (Group 1) and 23% with three classes per grade (Group 2). For both groups we use graphical and statistical tests to examine out student sorting evidence based on previous school marks.

In graphical terms we could not identify any evidence of non-random assignment in the Chilean school system. Statistical tests were needed to investigate the hypothesis more formally. We designed 12 *Non-RA* Indexes per school, where we compare real classes with artificially created counterfactuals. All possible comparisons depend on the type of counterfactual used (RDM, SRT), the statistical test (**t-test**, **ks-test**), and the measure performance (GPA, Language, Maths).

The *Non-RA* Index provides a measure of SoE per school, and depending on its score we classify schools into five groups of non-random assignment evidence: *None, Low, Medium, Med-High* and *High*. For every *Non-RA* Index, we aggregate the school classification by group of schools (**Group 1** and **2**), as in principal the requirements to classify schools into each category slightly differs because schools from **Group 2** have more possible comparisons.

The results for **Group 1** suggest that in average, 69% of schools do not present any evidence of non-random assignment, while 25% is classified with *Low* SoE, and only 5% shows *Medium* levels of non-random assignment.

Considering schools in **Group 2**, we suspect there are higher chances of practising non-random assignment of pupils to classrooms, as having one class more increases the option for grouping students. The results presented in Tables 4.23 and 4.24 confirm that more schools have *Low* evidence of non-random assignment. Compared to **Group 1**, the percentage of schools classified with *None* SoE decreased from 69% to 46%, while the concentration of schools in the *Low* level of SoE increased from 25% to 42%.

However, the total decrease from **Group 1** to **Group 2** in the proportion of schools classified with *None* or *Low* levels of SoE was approximately 6%. If we aggregate both school groups observations and calculate the percentage average of schools in each category (see Table 4.25) we observe that 64% of schools do not show any evidence of non-random assignment of pupils to teacher or classroom. Schools without SoE or *Low* evidence of it, represent in average the 93% of our whole sample for Cohort 1.

It is worth noting, that the average of schools with *None* SoE would improve for all *Non-RA* Indexes if we consider significance level of 10% instead of 5% when we are testing fort the mean and the distribution differences between real classes and their counterfactuals. We interpret the results presented in this chapter as strong enough to suggest a very low evidence of non-random assignment of pupils to classes in the Chilean school system, at least between 4^{th} and 5^{th} grades.

Chapter 5

Teacher effects estimations in the Chilean primary school system

Abstract

In this chapter, we present the estimation methodology and model specification used to identify teacher unobserved effects on pupil performance. We show "Teacher Effects" (TEs) estimates in terms of standard deviation (SD) of the empirical Bayes (EB) distribution, and we discuss their predictions, measuring the expected impact of being exposed to higher effective teaching.

To our knowledge, the thesis contains the first teacher effect estimates for 4^{th} grade general teachers based on the very rich dataset corresponding to the Chilean school system. We find that if a student from the 50^{th} percentile of the National examination distribution is taught by a 1 SD more effective teacher, she would move up 9 percentile ranking positions in Language and 12 in Maths.

We also test for heterogeneity of teacher effects, and we find that teacher effectiveness in Municipal schools is more heterogeneous compared with teachers in private schools. In general, our results indicate a substantial contribution of 4^{th} grade general to pupils' academic performance within the Chilean school system, in addition to the teacher and school observable characteristics. The evidence of non-random assignment of student to teachers found in Chapter 4, and the interesting findings shown in this chapter, support the recommendation of implementing our TE estimates as a complementary instrument for the National teachers evaluation in the Chilean primary schools.

Contents

5.1 Introduction

In Chapter 2, we presented a detailed review of how "Teacher Effects" (TEs) estimates can be obtained from Value Added Models (VAMs) typically found in the literature. Starting from a cumulative education production function, we derived a general VAM specification, in which we imposed different parameter restrictions to construct four different models. We explained the most common approaches used to estimate TEs, discussing the crucial assumptions that were necessary to ensure consistency of estimators.

For the purpose of our own analysis, we choose the VAM that enables us to predict TEs, separately from unobserved student ability (SA) and school effects (SEs). To estimate these effects simultaneously we need to set some assumptions and restrict the model specification, besides selecting a particular sample from the original dataset.

We estimate our proposed VAM using Maximum Likelihood estimation (MLE) methods and obtain predictions of individual, teacher and school effects from the estimated empirical Bayes (EB) distributions.[1] We discuss the assumptions that we need to consistently estimate our model using this method. For example, we require a strict exogeneity assumption for some covariates, non-random student to teacher assignment, and some additional independence conditions for unobserved heterogeneities.[2]

The main focus of this chapter is the TEs estimation and the parameter of interest is the standard deviation (SD) of the EB distribution of TEs. To interpret intuitively the SD of TEs, we ask the following question: given the actual percentile ranking of a student in the sorted scores distribution, what is the effect of a "higher" quality teacher on this particular student? We propose a hypothetical treatment to predict the impact. We take, as a reference, an average pupil from the mean or the median of the examination scores distribution within

[1] See Chapter 2, subsection 2.3.3 a detailed description of the empirical Bayes approach.

[2] Note, this estimation method can be thought of as a Quasi-MLE method when normality assumption for the unobservables are not made.

the sample, and we analyse the expected change in her percentile ranking when she is exposed to the treatment. The treatment consists of exposing the student to a 1 SD more effective teacher, *ceteris paribus*.[3]

Our results suggest, that general teachers of 4^{th} grade are more able to generate larger impacts on Maths than Language. If a pupil from the median of the distribution of The National Exam (Simce) scores were exposed to 1 SD more effective teacher, she will move up around 9 percentiles positions in Language and 12 in Maths.[4]

These findings confirm the substantial importance of teachers on academic performance, independent from their observable characteristics. Teacher ability can make a big differences in pupil performance, even after controlling for observed and unobserved student, class, and school factors.

With respect to the estimated coefficients, we find a significant effect of gender on pupil achievement with girls outperforming boys in Language but boys performing better than girls in Maths. The estimated gap in Maths performance motivates us to analyse whether there are differences in TE estimates in single-sex schools. Therefore, we select another sub-sample of single-sex schools to estimate the same VAM by MLE and obtain the TEs with the EB method (MLE-EB).

Estimated SD of TEs in girls-only schools seem to be the same as the estimates obtained from all schools, but the predicted movement on pupil ranking in Language is larger from girls-only schools. Nevertheless, the expected impact on Maths achievement is lower in this type of school. With respect to estimated TEs in boys-only schools, we can infer the point estimate of this impact on Maths and Language is lower compared to the point estimates for the whole sample.

However, both types of single-sex schools represent small samples with respect to the original selected cohort, with just 6% and 3% of the original selected cohort distributed in girls-only and boys-only schools, respectively. It is important to highlight that the small size of these samples could lead us to non-robust conclusions, particularly for boys-only schools, which only considers 36 establishments and 86 teachers. An additional issue one has to bear in mind here is the possibility of sample selection biases due to parents choosing to send their child to a single-sex school.

Finally, to check for heterogeneous effects, we estimate TEs with and without considering private schools, arguing that private schools are more likely to select students and teachers, violating random assignment assumptions. Hence,

[3]The analysis is based on the assumption that no one else within the distribution is being affected by the hypothetical treatment.

[4]We define 100 positions, so the predicted movements from the median are with respect to the student in the 50^{th} position, while the movements from the student in the mean will depend on the specific distribution.

we test whether the potential non-random assignment would generate some bias in our estimator. Firstly, we find no difference in expected teacher impacts on pupil performance when *Unsubsidised Private* schools are removed from the sample. Secondly, we find that teachers from Municipal schools are more heterogeneous, having potentially larger impact on pupil achievement. Moreover, the predicted impacts do not differ when we estimate *Municipal* school teacher effects with and without private school observations.

These results suggest the existence of heterogeneous teacher effects, and they can be equally estimated either with or without including private schools in the sample. However, the similarity in the estimation results using different samples by type of school, suggest that potential differences in non-random assignments of pupils to schools and teachers to schools seem not to affect the TE estimates in the Chilean school system.

5.2 VAM: Specification and Estimation

The composition of an educational system takes the form of a hierarchical organisation. Students are nested into classrooms which are taught by a particular set of teachers. These teachers can teach in different grades, classrooms, and sometimes even in more than one school. Students cannot be enrolled simultaneously in more than one grade or school. In those cases where students are not exclusively nested into teacher-classroom or teachers into grades-schools, the complexity of TE estimation increases considerably, particularly from the computational point of view.

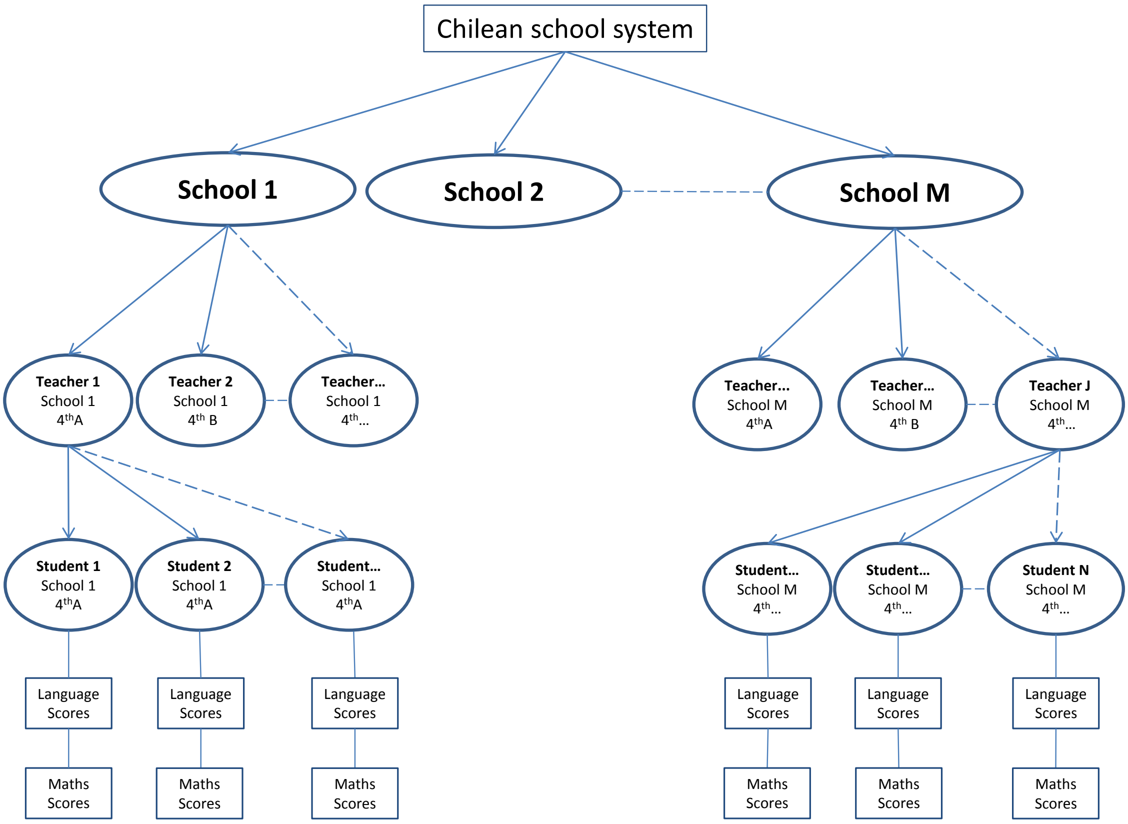
Given the difficulties of TE estimations when students follow specific-subject tracks with specialised teachers per area (e.g. Language, Maths or Science), most studies focus on Value Added estimation for primary school teachers. In early primary grades, from 1^{st} to 4^{th} grade, it is usual to find classes taught by general teachers, in contrast to late primary grades (from 5^{th} to 8^{th}), where specialised subject (SS) teachers are more commonly observed.[5]

Some of the related studies in the literature concentrate their analysis between 3^{rd} and 6^{th} grade cohorts (e.g. Rockoff (2004); Rothstein (2009, 2010); Chetty et al. (2014a)).[6] For the Chilean educational context, we are also interested in cohorts within this range, choosing 4^{th} grade cohorts because of the availability of the National Simce exam scores, and the higher proportion of general teachers compared to SS teachers.[7]

Figure 5.1 shows what we need in order to identify, individual ability, teacher effectiveness and school quality within our VAM framework. We are able

[5]General teachers are allocated to a single classroom and they teach all main subjects, or at

Figure 5.1: Identification strategy



to distinguish teacher effects from school effects only in schools with at least two classes per grade, implying observing at least two general teachers per school. Hence, Figure 5.1 represents the M schools which satisfy the school sample selection, joined with the J teachers and N students who also satisfy the selection criteria. It is important to highlight the fact that we are not able to separate teacher and classroom effects, as general teachers are assigned to one classroom per year.[8] Also note, the use of Maths as well as Language scores separately, enables us to identify the effect of unobserved *ability* of the student.

In Chapter 2, we discussed four different VAM specifications, their associated parameter restrictions, and assumptions required to consistently estimate the parameters and TEs using different estimation techniques. In the current Chapter, we use a modified version of **Model 4** (equation (12) of Chapter 2) to estimate TEs and SEs by explicitly including the total school contribution $S_g'\theta_g + s_g$.

least Language and Maths.

[6]See Appendix 2.1, the summary of datasets found in the VAM Literature.

[7]The timeline of when the National standardised examination (Simce) is taken is provided in Chapter 3. The percentage of general teachers in 3^{rd}, 4^{th} and 5^{th} during 2004-2010, for three specific cohorts, is presented in Chapter 4.

[8]Other studies such as Chetty et al. (2014a) cannot separate teacher effects from specific classroom effects in early primary grades.

$$A_{i,g} - \lambda_0 A_{i,g-1} = x'_{i,g}\beta_g + T'_{j,g}\pi_g + \tau_{j,g} + S'_g\theta_g + s_g + \alpha_i + \varepsilon_{i,g} \tag{5.1}$$

where λ_0 $(0 < \lambda_0 \leq 1)$ is the preset persistence parameter. This VAM uses the student's academic performance (Language and Maths scores) as response variable $A_{i,g}$, and we have set S_g and $T_{j,g}$ to be the observed school and teacher characteristics, while s_g and $\tau_{j,g}$ are the unobserved parts. The sub-index j identifies the teacher, and g the grade in which student i is enrolled.[9]

The theoretical model, based on the General Achievement Function (GAF), aggregates all educational inputs into the vector e_i as we showed earlier in equation (2.8). We make the difference between the total contribution of schools and teachers with their specific unobserved effects. Hence, our empirical approach includes observed school and teacher characteristics in the estimation and our TE (and SE) estimates are controlling for the means.

In the literature we have found different definitions of teacher effects, and whether to include or not teacher observable characteristics depends also on the estimation strategy and the availability of data at individual teacher level.

Authors assuming teacher fixed effects do not include teacher observable characteristics into their VAM specifications (Ballou et al. (2004); Aaronson et al. (2007)). Other authors consider all teacher characteristics (observed and unobserved) as overall teacher random effects (Sanders and Horn (1994); McCaffrey et al. (2004)), while Chetty et al. (2014a) separate in time-varying and time-invariant characteristics. On the other hand, there are authors such as Rockoff (2004) and Jacob and Lefgren (2008) who control for teacher experience and classroom level covariates, respectively. We follow the last two studies including teacher, classroom and school level characteristics.

Therefore, our VAM specification presented in equation (5.1) is characterised by a multilevel structure. To estimate teacher effects, we follow the "Random Effects"(RE) Approach (or **Approach 3** from Chapter 2) and estimate the model using the Maximum Likelihood estimator (MLE-EB) where the teacher effects τ_j, the school effects s_g, and the individual ability α_i are all specified as random effects. We obtain estimates of these effects using the mean of the posterior empirical Bayes (EB) distributions.

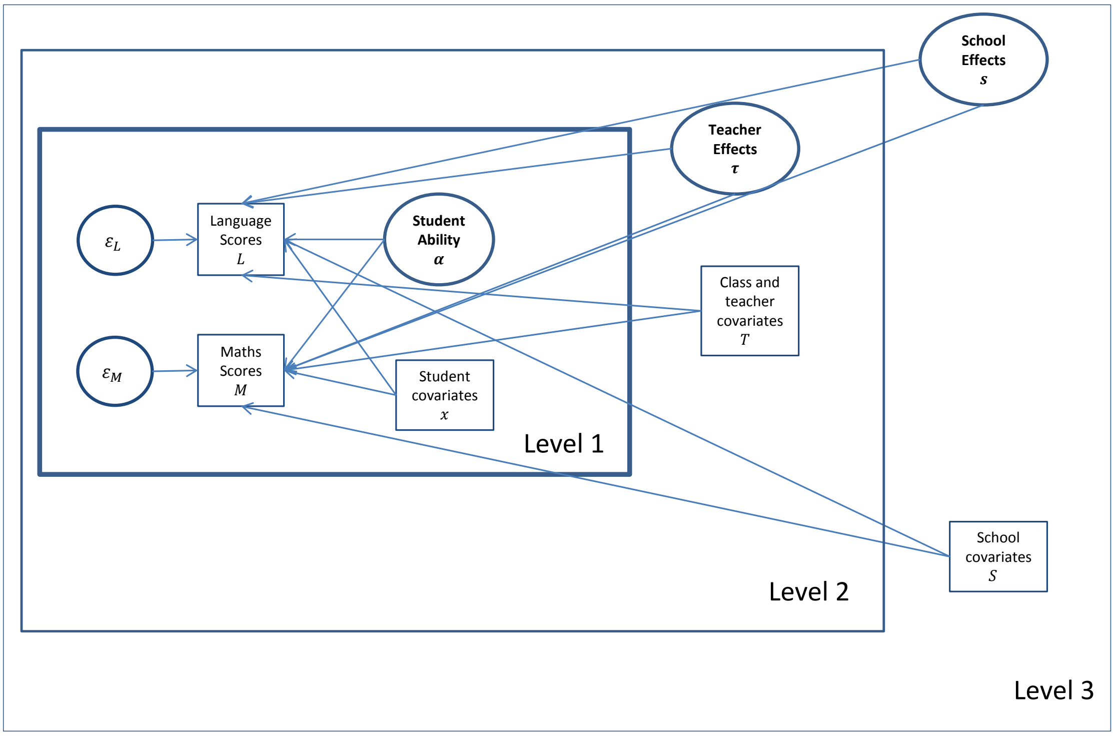
It has been suggested that under random assignment of teacher to classroom, and particularly for small cross-sectional samples, the MLE-EB methodol-

[9] We have suppressed schools and subjects sub-index for simplicity of the expression. Given the cross-sectional nature of our model, the overall equation error term is not written as the MA(1) error as derived in Chapter 2.

ogy performs well in estimating TEs (Guarino et al. (2014a,b)).[10]

To make the estimation of equation (5.1) more comprehensible, we show the model in a path diagram representation. Here we separate the explained variable $A_{i,g}$ into two response variables for pupil achievement in both Language $L_{i,g}$ and Maths $M_{i,g}$. Figure 5.2 contains both boxes and circles, where boxes correspond to observed variables (x, T, S) and circles unobserved heterogeneities (α_i, τ, s). The arrows show the linear relationships between them, and the direction states their causality.

Figure 5.2: Path diagram for teacher effects estimation



All arrows in Figure 5.2 go to the response variables (L, M), including circles from the error terms $\varepsilon_M, \varepsilon_L$. The large border boxes show how observed and unobserved variables are aggregated or clustered at three different levels (individual, teacher-classroom, and school level). Thus, the path diagram is an alternative form with which to present equation (5.1) as a simultaneous equation system that explain Language and Maths performance of students.

To facilitate the discussion of the assumptions, we first rewrite (5.1) separately for Language and Maths:

$$L_{i,4} - \lambda_0 L_{i,3} = x_i'\beta_L + T_j'\pi_L + S'\theta_M + 1.\alpha_i + \delta_{L,\tau}\tau_j + \delta_{L,s}s_g + \varepsilon_{i,L} \tag{5.2}$$

$$M_{i,4} - \lambda_0 M_{i,3} = x_i'\beta_M + T_j'\pi_M + S'\theta_M + \delta_{M,\alpha}\alpha_i + 1.\tau_j + 1.s_g + \varepsilon_{i,M} \tag{5.3}$$

[10] As discussed in detail in Chapter 4, we did not find strong evidence of non- random assignment of students to classes/teachers in our sample.

where the subscripts 4 and 3 refer to the grades. M and L are the test scores for Maths and Language exams, respectively. The observable student/family, class/teacher and school characteristics are collected in vectors x, T and S respectively. All these variables are assumed to have subject-specific effects. There are three unobservable random effects at the pupil, teacher and school levels: α_i, τ_j, s_g. The correlation between the pair of same random effects in the two equations are captured using a factor loading representation as shown above. For identification we impose the following normalisation restrictions, $\delta_{M,\tau} = \delta_{M,s} = \delta_{L,\alpha} = 1$.[11] The other three factor loadings ($\delta_{M,\alpha}, \delta_{L,\tau}$ and $\delta_{L,s}$) are freely estimated.

We now turn to the assumptions required to obtain consistent estimators of (5.2) and (5.3) using the above MLE-EB method. The most important assumption for all VAM estimators is ***Assumption A1*** which is the strict exogeneity of covariates conditional on past, present and future values of the observables and unobservables.[12]

Assumption A1:

$E[\varepsilon_{i,M}, \varepsilon_{i,L} | past,\ present$ and $future$ values of $(x, T, S), \alpha_i, \tau_j, s_g] = 0.$

The above strict exogeneity assumption will fail, if for example the students are sorted into classes based on either observable or unobservable characteristics or both. This leads us to the second set of assumptions we require; ***Assumptions A2.1***, ***A2.2*** and ***A2.3***.

Assumption A2.1: Random assignment of students to schools.

Assumption A2.2: Random assignment of students to teachers (or teacher to classrooms).

Assumption A2.3: Random assignment of teachers to schools.

There are three potential sources of non-random assignment that can lead to endogeneity of covariates: (i) **student-to-school**, (iii) **student-to-teacher** and (iii) **teacher-to-school**. In Chapter 2, we discussed how the failure of these assumptions can lead to inconsistent estimators.

Regarding the first source of endogeneity, there is little we can do beyond

[11]The restriction where one of the path coefficients from a latent variable is set equal to 1 is also known as the normalisation constraint or as the unit-loading rule (Rabe-Hesketh et al. (2004); Heckman et al. (2006)). The *loading factor* which is constrained to 1 is also called the *anchor* value for the corresponding latent variable. The inference made on the estimated *loading factor* is with respect to the normalised factor.

[12]We suppress the school sub-index for simplicity.

controlling for family background and student characteristics, including previous academic performance (Rothstein (2009, 2010); Chetty et al. (2014a)). In all non-experimental contexts there is always a degree of endogeneity in student to school assignment, as parents either choose schools for their children or choose where to live, which might be also related to school decisions.

We address the second source of endogeneity based on the results obtained in Chapter 4. We assume there is at most only weak evidence of students being sorted into classrooms with respect to previous school marks. [13]

In relation to the third source of endogeneity, like most papers in the literature, we are unable to demonstrate random teacher-to-school assignments. We have to assume random allocation of teacher to schools. However, we check whether estimated teacher effects differ by different types of school.

The next set of assumptions deal with the distribution of the random components.

Assumption A3.1: $E[\alpha_i | past, present$ and $future$ values of $x] = 0$.

Assumption A3.2: $E[\tau_j | past, present$ and $future$ values of $T] = 0$.

Assumption A3.3: $E[s | past, present$ and $future$ values of $S] = 0$.

Note, under ***Assumption A2.1*** to***A2.3***, the requirements for ***Assumptions A3.1*** to ***Assumptions A3.3*** have become simpler. For example, if there is random sorting of students to teachers/classes and also of teachers to schools,

$$E[\alpha_i | past,\, present \,\text{and}\, future\,(x, T, S)] = E[\alpha_i | past,\, present \,\text{and}\, future\, x]$$

Given the above assumptions, we need some distributional assumptions in order to implement MLE-EB. We assume that all the random errors are *independent of each other*, and have the following distributions:[14]

$\alpha_i \sim iid\ N(0, \sigma^2_\alpha)$; $\tau_j \sim iid\ N(0, \sigma^2_\tau)$; $s_g \sim iid\ N(0, \sigma^2_s)$; $\varepsilon_{i,M} \sim iid\ N(0, \sigma^2_\varepsilon)$; $\varepsilon_{i,L} \sim iid\ N(0, \sigma^2_\varepsilon)$.

[13] In Chapter 4, we showed there is not enough evidence of non-random assignment of student-to-classrooms (or teachers). The analysis was made for the Chilean school system with three 4^{th} grade cohorts using a selected sample of schools, teacher and students.

[14] The assumption that $\varepsilon_{i,M}$ and $\varepsilon_{i,L}$ are *independent of each other* is similar to the independence assumption proposed by Kane et al. (2008) who suggests there is a non-persistent component of error which is independent from class to class, and in our case it would be independency from Language to Maths.

Note, we restrict the variances to be the same for $\varepsilon_{i,M}$ and $\varepsilon_{i,L}$.[15]

Regarding the estimation, we first estimate equations (5.2) and (5.3) simultaneously using MLE under the assumption that all error terms are normally distributed. In the second stage, we estimate the expected value of the posterior distributions obtaining the EB prediction or estimation.[16] If the above assumptions hold, MLE-EB will provide consistent estimators.

In summary, we conduct our analyses using the National standardised test scores of Simce exam for grade 4, and the standardised school marks as our measures of achievement in 3^{rd} grade for our VAM. We also deviate from the literature and identify the student level, teacher level and school level effects. Given the cross-sectional nature of our model, we do this by using: (i) the Maths and language test scores separately (instead of just one score); (ii) a sample of schools that had at least two classes in the 4^{th} grade with general teachers teaching both subjects.

Next section describes in detail the sample selection process used in order to estimate the above model and identify student ability α, teacher effects τ, and school effects s from our VAM estimation.

5.3 Sample selection

To identify student and school effect distributions, it is necessary to select a sample from the original 4^{th} grade 2005 cohort. The sample selection applied is necessary to estimate simultaneously the equations (5.2) and (5.3) by MLE.

We propose to estimate TEs based on general teachers, and as we have seen in Chapter 4, these teachers are mainly concentrated in primary grades from 1^{st} to 4^{th} grade. Teachers are more specialised by subject from 5^{th} grade onwards, where they also teach to more than one class or grade per year.[17]

[15] We could not achieve convergence in the estimation of the models with unrestricted error variances for $\varepsilon_{i,M}$ and $\varepsilon_{i,L}$ and hence the models are estimated under the restriction that the variances are the same. Also note, the dependent variable was standardised with respect to the *National* Simce distribution mean and variance. Due to the particular selection of the sample, the mean and variance of the Simce score in the estimation sample are not 0 and 1 respectively. However, they are similar. If the standardisation were carried out with respect to our estimation sample values, we would generally expect the $\varepsilon_{i,M}$ and $\varepsilon_{i,L}$ to have variances which are similar and close to one.

[16] To estimate the model and obtain the EB estimates for the latent variables, we use the Generalised Structural Equation Modelling (GSEM) programme created by Rabe-Hesketh et al. (2004, 2007) and available in Stata 13 (StataCorp, 2013).

[17] In Chapter 4, from Table 5 and 6, we observe how the distribution of specialised subject (SS) teachers is for 4^{th} grade cohorts in 2005, 2007 and 2009, considering their 3^{rd} and 5^{th} grade respective observations. In this case, we observe that the percentage of specialised teachers increases over 67% of classes from 5^{th} grade, and it has been gradually growing in 4^{th} grade classes since 2005, getting up to 25% of the 4^{th} grade classes in 2009. Therefore, we chose the 4^{th} grade 2005 cohort to estimate initial TE, where the number of classes with SS teacher just

We also have shown that at primary level, the Simce have been taken only in 4^{th} and 8^{th} grade, at least until 2013. Since we need current and lagged test scores to estimate our model given by equations (5.2) and (5.3), we choose the 4^{th} grade 2005 cohort for which we recover school marks of 3^{rd} grade 2004 to use as a proxy for lagged scores.[18]

The selected sample satisfies a series of conditions that we describe in Table 5.1. Here, we show the order in which we eliminate observations that did not fulfil our criteria. We classify the selection process into three groups: (i) selection related to teacher conditions, (ii) selection with respect to availability of information at student level, and (iii) selection by school level conditions.

The selection criteria ensure that we have general teachers correctly assigned to one specific class. Thus, we drop classrooms without an assigned teacher, and whose teachers were teaching in more than one class (i.e. SS teachers) or teaching in more than one school (potentially mover teachers).

As Simce scores, school marks and observable characteristics are essential for the VAM estimation, we eliminate observations without this information available. Repeating students from the 4^{th} grade 2004 cohort are excluded from the sample because we want to distinguish them from those studying 4^{th} grade contents for the first time.

Following the literature, we also introduce restrictions on the number of students per class in order to reduce potential bias in estimation due to few observations per teacher (less than 15 students per grade).[19] We also eliminate students belonging to classes with more than 45 students enrolled because this number might lead to misclassification problems as 45 students is the maximum number of students per classroom allowed by law.

To understand the weight of every type of selection in terms of students, teachers and schools dropped, we construct Table 5.2. In this table, we observe the number of students, teachers and schools removed from the original cohort as a result of the sample selections.

represents 14% of total classes.

[18] In Chapter 3, we checked the correlation between standardised Simce scores and standardised school marks to support the use 3^{rd} grade school marks (standardised at school level) as a proxy for unobserved National examination scores, unavailable for 3^{rd} grades.

[19] See Appendix 2.1, the summary of datasets in the VAM literature, where we show some of the selection criteria used by authors and it can be seen those which coincide with ours.

Table 5.1: Sample Selection Criteria - 2005 4^{th} grade cohort

Selection Sample		Criteria	Dropped obs.	% Dropped
	Total observations original bases:		**268,162**	**100%**
	(1) Dropping obs. without teacher ID assigned	Yes	3,706	1%
Teacher selection	(2) Dropping obs. with teachers observed in 2 o more schools	Yes	23,973	9%
	(3) Dropping obs. with specialised teachers	**Yes**	**37,458**	**14%**
	(4) Dropping obs. without teachers characteristics	Yes	14,671	5%
	(5) Dropping obs. without Simce Exam	Yes	13,084	5%
Student selection	(6) Dropping obs. with repeating students (Not observed in 3rd grade 2004)	Yes	8,380	3%
	(7) Dropping obs. without School Marks - 4th grade 2005	Yes	301	0%
	(8) Dropping obs. without School Marks - 3rd grade 2004	Yes	1,011	0%
	(9) Dropping obs. without student characteristics	**Yes**	**19,615**	**7%**
	(10) Dropping obs. without school characteristics	Yes	-	0%
	(11) Dropping obs. with class letter "F" label	Yes	305	0%
School selection	(12) Dropping classes without a minimum No. of students per class	15	13,431	5%
	(13) Dropping classes without a maximum No. of students per class	45	-	0%
	(14) Dropping schools without a minimum of classes per grade	**2 or more**	**41,592**	**16%**
	(15) Dropping schools with less general teachers than classes per grade	Yes	925	0%
	Total dropped observations		**178,452**	**67%**
	Total observations selected sample:		**89,710**	**33%**

Notes: **(i)** We dropped observations sequentially from the original base "4^{th} Grade -2005" cohort. **(ii)** The first part of the selection shows the number of observations eliminated on teacher requirements. **(iii)** The second part drops observations on student conditions. **(iv)** The third part we drops observations related to school, grade and classroom conditions. **(v)** We eliminate pupils attending classrooms labeled as F in row (11) because this was our classification to all classes when the letter assigned in the original database was not between A and E, and it might represent an error in the code.

The original 4^{th} grade 2005 cohort is composed of 268,162 students, 12,233 teachers (including general and specialised teachers) and 8,338 schools. After the selection process, we have a cohort with 89,710 students, 1,337 schools and 3,151 general teachers. The most important source of selection is coming from dropping observations of schools with less than two classrooms per grade (row 14, Table 5.1). The second largest source of selection is at teacher level, when observations are eliminated because teacher are specialised in subject (row 3, Table 5.1).

Table 5.2: Number of observations dropped by type of selection

Original 4th grade cohort		**Before selection**	**%**
	Number of Students	268,162	100%
	Number of Teachers	12,233	100%
	Number of Schools	8,338	100%
		Dropped observations	**%**
Teacher selection	Number of Students	79,808	30%
	Number of Teachers	4,857	40%
	Number of Schools	2,959	35%
		Dropped observations	**%**
Student selection	Number of Students	42,391	16%
	Number of Teachers	555	5%
	Number of Schools	545	7%
		Dropped observations	**%**
School selection	Number of Students	56,253	21%
	Number of Teachers	3,670	30%
	Number of Schools	3,497	42%
Selected 4th grade cohort		**After selection**	**%**
	Number of Students	89,710	33%
	Number of Teachers	3,151	26%
	Number of Schools	1,337	16%

Note: (i) From the original 4th grade 2005 cohort, the sample selection is carried on sequentially, through teacher, student and school conditions, and the Tables show the number of observations dropped at each level. (ii) The percentage presented in the last column indicate the proportion of students, teachers or schools with respect to the original cohort.

The descriptive statistics for this particular cohort, before and after applying the selection process, is shown in Table 5.3. Comparing the original 4^{th} grade 2005 and the selected group of schools, we can infer from Table 5.3 that there are no considerable differences between them in terms of pupil academic performance, such as GPA, Language and Maths school marks and Simce scores. The selected sample has slightly higher average in every achievement measure for 2004 and 2005. Note that, as we mentioned earlier, the mean of the standardised Simce scores shifted from 0 to 0.1, and the standard deviation decreased from 1 to 0.98. Despite these small differences, it seems that the selected sample is a good representation of the original cohort.

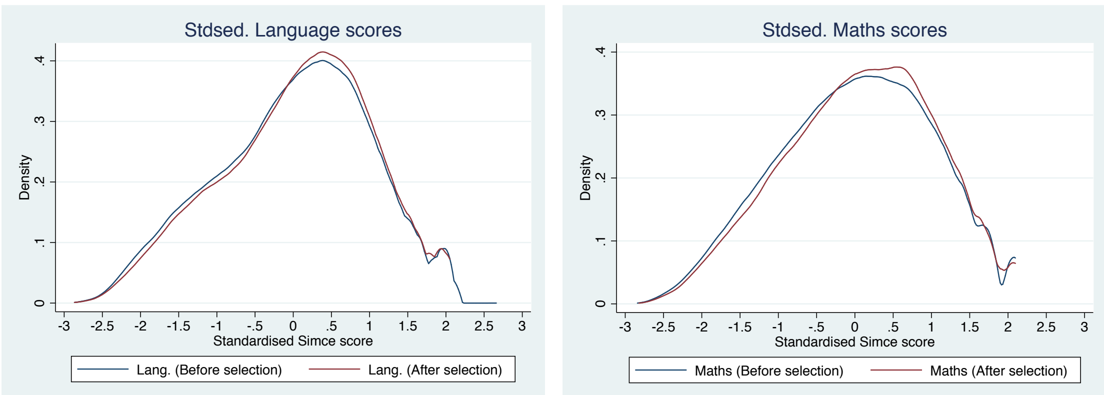
Table 5.3: Descriptive statistics - Before and after selection
4^{th} grade cohort 2005

Before selection	Mean	Std. Dev.	Min	Max
Pupil Level				
GPA 4th 2005	5.7	0.9	1	7
GPA 3th 2004	5.9	0.8	1	7
School Language Marks 2005	5.4	0.8	1	7
School Maths Marks 2005	5.3	0.9	1	7
School Language Marks 2004	5.6	0.8	1	7
School Maths Marks 2004	5.5	0.9	1	7
Language Simce Scores 4th 2005	255.6	53.2	103.1	397.2
Maths Simce Scores 4th 2005	247.7	55.3	90.7	363.6
Stdsed. Lang. Simce Score 4th 2005	0.0	1.0	-2.9	2.7
Stdsed. Maths Simce Score 4th 2005	0.0	1.0	-2.8	2.1
Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.2	0.6	8	12
Attendance	92.5	14.0	0	100
Speecial Needs	0.0	0.1	0	1
Mother Education	2.8	2.1	0	8
Father Education	3.1	2.4	0	8
Household Income	2.2	2.7	0	12
Class Level				
Class size	23.3	14.6	1	54
Peers Average GPA	5.7	0.5	1.0	7.0
Teacher Level				
Gender (Female=1)	0.8	0.4	0	1
Years of experience in the system	22.8	13.2	0	40
(Teaching hrs / Contract hrs) Ratio	0.9	0.2	0.02	1
School Level				
Municipal Schools	0.6	0.5	0	1
Private Voucher Schools	0.3	0.5	0	1
Unsubsidised Private Schools	0.1	0.2	0	1
Rural Area	0.5	0.5	0	1
Number of classes per grade	1.4	0.8	1	12
Number of students per grade	32.2	37.6	1	535
Number of students	268,162			
Number of teachers	12,233			
Number of schools	8,338			

After selection	Mean	Std. Dev.	Min	Max
Pupil Level				
GPA 4th 2005	5.9	0.6	1	7
GPA 3th 2004	6.0	0.8	1	7
School Language Marks 2005	5.5	0.8	1	7
School Maths Marks 2005	5.4	0.9	1	7
School Language Marks 2004	5.7	0.8	1	7
School Maths Marks 2004	5.6	0.9	1	7
Language Simce Scores 4th 2005	258.5	52.3	103.1	364.8
Maths Simce Scores 4th 2005	251.1	54.2	91.7	363.6
Stdsed. Lang. Simce Score 4th 2005	0.1	0.98	-2.9	2.1
Stdsed. Maths Simce Score 4th 2005	0.1	0.98	-2.8	2.1
Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.2	0.4	8	12
Attendance	94.4	6.0	0	100
Speecial Needs	0.0	0.1	0	1
Mother Education	2.9	2.0	0	8
Father Education	3.1	2.2	0	8
Household Income	2.2	2.6	0	12
Class Level				
Class size	28.5	6.8	15	45
Peers Average GPA	5.7	0.4	2.3	6.7
Teacher Level				
Gender (Female=1)	0.9	0.3	0	1
Years of experience in the system	20.0	11.5	0	40
(Teaching hrs / Contract hrs) Ratio	0.9	0.1	0.11	1
School Level				
Municipal Schools	0.6	0.5	0	1
Private Voucher Schools	0.4	0.5	0	1
Unsubsidised Private Schools	0.0	0.2	0	1
Rural Area	0.0	0.2	0	1
Number of classes per grade	2.4	0.6	2	5
Number of students per grade	67.1	25.2	30	209
Number of students	89,710			
Number of teachers	3,151			
Number of schools	1,337			

Notes: (i) The selection sample was made for identification, and the most important sources of dropped observations were schools with less than two classes per grade and schools and schools with specialised teachers only, both sum up to 20% of the original pupil's observations in the cohort. **(ii)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(iii)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600.

Figure 5.3: Kernel distributions: Stdsed. Simce Scores - Before and After Selection
4^{th} grade cohort 2005



If we focus particularly on the standardised Simce scores before and after the sample selection, we observe in Figure 5.3 that the Language and Maths distributions after selection accumulate slightly more students between 0 and 1 standardised Simce scores. Although, even if the distribution shifted marginally to the right, their shape and standard deviation are still very similar.

Essentially, the main differences are driven by the selection of schools with at least two classes per grade and for which we were able to identify their 4^{th} grade general teachers. In Table 5.3, we observe how the largest differences are given by the average class size (23.3 versus 28.5), the number of classes per grade (1.4 versus 2.4) and the number students per grade (32.2 vs 67.1) in the original cohort and the selected sample, respectively.

5.4 Results

In this section, we present and compare the results obtained for three different estimations. As we discussed in Section 2, given our model assumptions, we have to assign a value for λ in equations (5.2) and (5.3) to obtain consistent estimators. In the first estimation exercise, we preset λ_0 equal to 0.4 (**Model 4**, Chapter 2) following evidence of estimated λ parameters in the literature. For the second estimation, we apply the commonly used full persistence assumption of previous input factors, as it is proposed in **Model 3**, from Chapter 2, setting the value $\lambda = 1$. The third estimation is carried on estimating λ within the model (**Model 1**, Chapter 2), but without treating the lagged test scores as endogenous.[20]

All three estimations consist of two-step procedures. In the first step, the VAM is estimated by MLE to obtain all the parameter estimates including the variances of the prior distributions of student, teacher, and school effects. In the second step, we employ the empirical Bayes (EB) approach to estimate the student ability (SA) α, teacher effects (TEs) τ and school effects (SEs) s as the mean of the posterior distribution for each unobserved heterogeneity.

Once the predicted individual, teacher and school effects are obtained, it is possible to determine their EB distributions, which will have smaller dispersions compared to the prior distributions.[21] Note the prior distributions of individual ability, teacher effects and school effects are assumed to be normally distributed with mean 0 and standard deviations equal to σ_α, σ_τ and σ_s respectively.

In Table 5.4, we show the estimated coefficients of the three VAMs, with

[20]Note that even if we are aware that the MLE-EB is inconsistent for Model 1, we estimate their TEs for comparison purposes.

[21]See Appendix 4.3, all empirical Bayes distribution tables of the three latent variables (α, τ, s), under the threes different settings of λ.

Table 5.4: Maximum Likelihood coefficient estimates
4^{th} grade 2005 selected sample cohort

	Lambda preset (λ=0.4)		*Lambda preset (λ=1)*		*Lambda estimated*	
	Model 4		**Model 3**		**Model 1**	
	Language	Maths	Language	Maths	Language	Maths
Loading factors						
School Effects	**1**	0.870***	**1**	1.141***	**1**	0.869***
		(0.015)		(0.035)		(0.014)
Teacher Effects	**1**	1.463***	**1**	0.985	**1**	1.521***
		(0.022)		(0.017)		(0.024)
Individual Ability	1.111***	**1**			1.134***	**1**
	(0.006)				(0.006)	
Student covariates						
Stdzd. Language School Marks 2004					0.518***	
					(0.002)	
Stdzd. Maths School Marks 2004						0.545***
						(0.002)
Gender (Female=1)	0.019***	-0.067***	-0.084***	-0.040***	0.005	-0.060***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Mother education level	0.032***	0.029***	-0.005***	-0.006***	0.027***	0.023***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Household income	0.019***	0.020***	-0.001	-0.002	0.016***	0.017***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Special needs	-0.351***	-0.284***	-0.059	0.002	-0.306***	-0.231***
	(0.052)	(0.050)	(0.049)	(0.050)	(0.051)	(0.049)
Class and Teacher covariates						
Class size	0.015***	0.016***	0.009***	0.009***	0.013***	0.014***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Peers average GPA	-0.520***	-0.618***	0.658***	0.681***	-0.336***	-0.389***
	(0.036)	(0.042)	(0.034)	(0.035)	(0.027)	(0.031)
Gender (Female=1)	0.079***	0.068***	0.046*	0.020	0.067***	0.054**
	(0.019)	(0.023)	(0.024)	(0.025)	(0.017)	(0.022)
Years of experience in the education system	0.004***	0.004***	0.005***	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Teaching ratio (Hrs teaching / Total hrs contract)	-0.068	-0.104*	-0.046	-0.092	-0.057	-0.096*
	(0.049)	(0.055)	(0.056)	(0.058)	(0.046)	(0.051)
School covariates						
Private voucher school	0.428***	0.426***	0.372***	0.350***	0.419***	0.414***
	(0.023)	(0.024)	(0.022)	(0.023)	(0.022)	(0.023)
Unsubsidised private school	1.136***	1.226***	0.973***	1.016***	1.100***	1.186***
	(0.052)	(0.053)	(0.047)	(0.050)	(0.049)	(0.049)
Rurality (Rural=1)	0.030	0.001	-0.050	-0.082	0.016	-0.015
	(0.054)	(0.055)	(0.051)	(0.054)	(0.052)	(0.053)
Constant	2.182***	2.830***	-4.460***	-4.510***	1.162***	1.541***
	(0.211)	(0.247)	(0.209)	(0.214)	(0.161)	(0.190)
Prior SD of School Effects: σ_s	0.318***	0.276***	0.228***	0.260***	0.310***	0.269***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Prior SD of Teacher Effects: σ_τ	0.239***	0.349***	0.342***	0.337***	0.210***	0.319***
	(0.004)	(0.004)	(0.006)	(0.006)	(0.003)	(0.003)
Prior SD of Individual Ability: σ_α	0.529***	0.476***			0.515***	0.454***
	(0.002)	(0.002)			(0.002)	(0.002)
Error variance: σ_ε^2	0.196***	0.196***	0.553***	0.553***	0.201***	0.201***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Log likelihood		-169784.8		-205338.9		-168767.1
Number of Observations		89,710		89,710		89,710

Notes: **(i)** To estimate the models and obtain the Posterior SD of unobserved heterogeneities, we use the Generalised Structural Equation Modelling (GSEM) programme, available in Stata 13 (StataCorp, 2013). **(ii)** Standard errors in parentheses. **(iii)** *** p<0.001; ** p<0.05; * p<0.01. **(iv)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(v)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600. **(vi)** Both Mother education level and Household income variables are assumed to be the same the year reported in the Simce Parents questionnaire and the year before, for which we do not have records. **(vii)** The base category for the included dummy variables is shown next in brackets: Pupil's gender (Male); Special needs pupils (all the rest pupils); Teacher's gender (Male); *Private voucher* schools (Municipal and Unsubsidised Private school); *Unsubsidised Private* school (Municipal and Private voucher schools); Rurality (Urban area).

and without the restrictions on the persistence parameter λ. The maximised value of the log-likelihood is the highest in the unrestricted model (**Model 1**) with the lowest value corresponding to the restricted model with $\lambda = 1$ (**Model 3**). In the following discussions, it should be noted that, given the model assumptions, estimator of **Model 1** may not be consistent if there is correlation between the lagged exam score and the error term ε.

At the top of Table 5.4 we find the *loading factors* estimates which indicate how much latent variables differ between Maths and Language in their impact on pupil attainment. The *anchor factor* (equal to 1) is taken as a reference or pivot. For example, in **Models 4** and **1**, we observe that the impact of school effects and individual ability are larger in Language than Maths, while the inverse occurs for teacher effects. The opposite results are obtained from the estimated *loading factors* in **Model 3** which does not include individual heterogeneity effects.[22]

The estimated effect of previous year's standardised school marks in **Model 1**, show that the estimated persistence parameter λ is very similar to 0.4 which we used in the estimation of **Model 4**, and is close to the range regularly observed in the literature.[23]

Not surprisingly, the signs and magnitudes of the estimated coefficients in **Model 3**, where we have set $\lambda = 1$, are different from the other two estimates. This model can be thought of as suffering from omitted variables problem given the omitted variable $A_{i,g-1}$ and its coefficient $(\lambda - 1)$, which disappear when $\lambda = 1$. Hence, the estimator is biased for the underlying true coefficients. Also note that, if $(\lambda - 1)$ is not equal to zero, α_i will not disappear from equation 2.6, in Chapter 2. This omission can cause additional biases. Therefore, we focus our discussions on the results obtained from **Model 4** and **Model 1**.

The first set of control variables corresponds to student covariates or variables at Level 1 (individual level). We see that among the individual covariates there is a negative and significant effect on Maths scores when the pupil is a girl. Given the estimated coefficient, we infer that girls would score between 0.06 and 0.07 SD lower in Maths compared to boys, ceteris paribus.

Mother's education and household income variables are categorical variables obtained from the parents' questionnaire, and they have been specially constructed to be compatible along all the student panel dataset (SPD).[24] The estimated coefficients indicate a positive and significant effects of these variables on both standardised Language and Math scores, with very similar levels of impact.

[22]See equation 2.6, derived from equation 2.3 in Chapter 2.

[23]Kane and Staiger (2008); Andrabi et al. (2011), have shown that the rate of persistence λ varies between 0.2 and 0.5. Hence, in the above estimations, we choose a λ_0 from the range $\lambda \in [0.4 - 0.5]$.

[24]See Chapter 3 for a full description of the Chilean Ministry of Education dataset.

Being entitled as a special need pupil has a significant and negative effect on academic performance; which is slightly larger in Language. Nevertheless, the number of pupils classified as special needs represents only 0.5% of the total selected sample. We have also estimated the VAMs excluding these students, but the coefficients do not change substantially.

Regarding class and teacher level characteristics, we find some interesting results. Unlike what it is commonly expected, the impact of class size seems to be positively correlated with pupils' achievement. Intuitively, the fewer the number of pupils taught per teacher the higher the expected outcome, and that is what some authors have found using experimental and quasi-experimental frameworks (Angrist and Lavy (1997); Krueger and Whitmore (2001)).

However, it is not clear yet whether the effect of class size on student performance has a negative relationship or whether the estimation is correct when is based on small size experiments and specific educational context. For example, Hoxby (2000) finds no significant impact of reduction in class size on pupils achievement.

Moreover, Urquiola and Verhoogen (2009) were the first to provide theoretical and empirical evidence of a non-linear relationship between class size and household income in the 4^{th} grade classes of the Chilean primary schools. The authors find an inverted-U relationship between class size and household income (and mother's education), predicting that high effective schools will have enrolment levels close to the size-cap (e.g. 45 per class), raising the tuition fees or increasing the pupils selection criteria rather than opening extra smaller classrooms. This situation may drive the positive sign of our class size coefficient.

We also find another counter-intuitive results with respect to peers average GPA. Nevertheless, we do not claim this is a peer effect estimation as we are aware of the potential bias involved in this type of estimation. On the contrary, this is a control variable which represents how far or how close the individual GPA is from the average of the class. As the class average GPA is lower than the student GPA, the higher the higher the achievement in the Simce exam.[25]

In relation to observable teacher variables, at the second level as well, we confirm there is a positive and significant effect in both subjects of being taught by a female general teacher than a male. Considering the case when $\lambda_0 = 0.4$, we observe that a pupil in the mean will have 0.08 and 0.07 SD higher attainment in Language and Maths standardised scores respectively, if the teacher is female

[25]The GPA is the average of final school marks in all subjects at the current period taken from the administrative *Performance data base*. It is reported by schools directly to the Ministry of Education. There is no a clear rule for which subjects are considered to calculate the individual GPA across schools. We decided not to standardised it and leave it in its original scale from 1 to 7.

instead of male. Muralidharan and Sheth (2015) find similar results, suggesting that female teachers are slightly more effective than males.

There is a very small effect of an additional year of experience in the education system on pupil attainment. This result is also in accordance with Rockoff (2004) findings, who states that teacher experience is only significant the first five years.

Additionally, there is a negative effect on student performance when increasing teaching hours over and above the total hours contracted, but the impact found is small and is not significant on Language scores, while the effect on Maths is only significant at the 10% significance level.

With the third level covariates, we control for type of school dependence and the rurality condition. From Table 5.4, we find that private schools (*Voucher* and *Unsubsidised*) have a positive and significant correlation with National scores. On the other hand, being rural does not have any significant impact on student performance.

In terms of sign, magnitude, and significance levels, the estimated coefficients presented in Table 5.4 are very similar under **Model 4** and **Model 1**. With respect to differences in impact between subjects (Language and Maths), all estimated coefficient signs move together apart from the female pupil identifier, where the effect of being a girl in Maths scores is negative while it is positive for Language scores.

We do not explore differences in academic performances by gender, but we propose that TEs might differ when teachers teach in single-sex environment. Thus, in the following subsection, where we present the TEs estimates for the whole sample, we also show the results obtained for single-sex school sub-samples.

5.4.1 Estimated Teacher Effects (TEs)

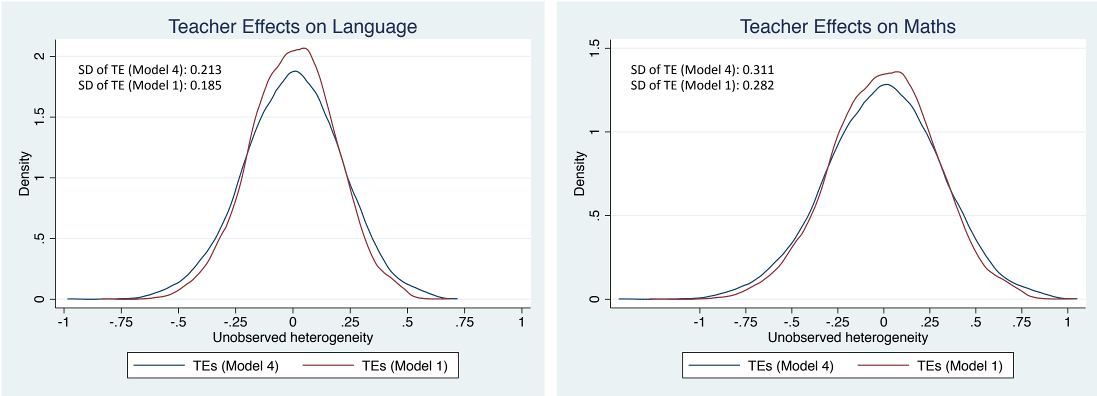
After the VAMs are estimated by the MLE, we obtain the EB distributions of student, teacher and school effects. However, to interpret the TE estimates more intuitively, we can either analyse their distributions graphically or illustrate their impact on potential changes in the percentile ranking of pupils' scores. Additionally, we also compare models' prediction in terms of teachers ranking correlation.

In Figure 5.4, we show the EB distributions of TEs obtained from **Models 4** and **1**.[26] Here, we observe how both VAMs predict similar distributions of TEs on both subjects, with differences of approximately 0.02 SD between them. However, Model 4 predicts more teachers in the tail areas relative to Model 1 which has

[26]The rest of the EB distribution plots for student ability and school effects are presented in Appendix 5.1 Their corresponding EB distribution tables are shown in Appendix 5.3.

more predicted TEs in the middle of the distribution.

Figure 5.4: Empirical Bayes distribution of Teacher Effects
4^{th} grade 2005 selected sample cohort



Regarding the analysis of estimated TEs based on their expected impact on pupil i's ranking, we first construct the ranking of students within our selected sample. We sort pupils into percentiles with respect to their standardised Simce score for Language and Maths, respectively. The distribution of pupils sorted by absolute and standardised Simce score is shown in 5.5.

Table 5.5: Simce scores distribution
4^{th} grade 2005 selected sample cohort

		All Schools	
Percentile		Language	Maths
10th	*Absolute Score*	183.4	176.3
	Standardised Score	-1.36	-1.29
25th	*Absolute Score*	222.8	213.0
	Standardised Score	-0.62	-0.63
50th	*Absolute Score*	263.8	254.2
	Standardised Score	0.15	0.12
75th	*Absolute Score*	296.5	291.3
	Standardised Score	0.77	0.79
90th	*Absolute Score*	323.5	320.1
	Standardised Score	1.28	1.31
Abs. Mean		258.5	251.1
Abs. SD		**52.3**	**54.2**
Stzd. Mean		0.054	0.062
Stdz. SD		**0.983**	**0.980**
Number of Schools			1,337
Number of Teachers			3,151
Number of Students			89,710

The scores for Language and Maths were standardised at National (4^{th} grade 2005 cohort) before sample selection, and Table 5.5 shows the Simce scores distribution of the selected sample cohort. In absolute terms, the mean and median of Language scores (258.5 and 263.8) are higher than Maths (251.1 and 254.2), but the dispersion (SD) is lower in Language than Maths (52.3 SD vs 54.2 SD). Thus, even if the estimated SD of TEs is the same for Language and Maths, the potential impact on the student i's when is exposed to 1 SD higher effective teacher would marginally differ in terms of ranking or percentile movement.[27]

Focusing on the SD of predicted TEs, we show in Table 5.6 (Part A), the mean and SD of TEs for Language and Maths. The difference between the SD of TEs in Maths and Language is close to 0.1 when λ is either equal to 0.4 or is estimated in the model. The fact that estimated TEs are higher in Maths than in Language, in both scenarios, is in accordance with what we have regularly found from the literature where higher SD of TEs are estimated for Maths than Language.[28]

In Table 5.6 (Part B), we present the impact of a hypothetical treatment which consist in exposing the pupil i, from the mean or the median of the Simce scores distribution, to 1 SD more effective teacher.[29] Then, we measure the predicted increase in her standardised Simce score and observe her movement within the ranking in terms of percentiles per subject.[30]

However, we are aware that the analysis using this hypothetical treatment is very restrictive as it requires strong assumptions to hold, such as no movements of other students within the same distributions or no other factors affecting changes. In practice, it is difficult to compare changes in teacher effects in terms of expected jumps. Thus, it is challenging to design and evaluate correctly, educational policies regarding the improvement of teacher effectiveness. We suggest, to consider our results as an interpretation of differences in predicted impacts between the subjects and the models employed.

The first case of 5.6 (Part B) is based on the pupil at the mean of the Simce scores distribution, whose ranking within the selected cohort is the 46^{th} percentile for Language and the 48^{th} for Maths. Under **Model 4**, when she is exposed to 1 SD more effective teacher, she would move to the 55^{th} and 60^{th} percentile ranking in Language and Maths, respectively. While using **Model 1** TEs estimations, the expected movement in the ranking would be to 54^{th} and 59^{th}, respectively. The difference between **Model 4** and **1** on the predicted impact of being treated hypothetically is just one percentile position.

The second case of 5.6 (Part B) is for the pupil in the median, whose corresponding percentile ranking is the 50^{th}. If the pupil were taught by a 1 SD

[27]See Appendix 5.3 the tables of Simce Score distribution in absolute and standardised scores for the original sample and the rest of subsamples used along this Chapter.

[28]See Chapter 2, Section 5: *"Impact of teacher effects: Standard deviation (SD) of measures"*.

[29]Since the estimated TEs refer to the effect on the conditional mean score, we only report the effect of this hypothetical treatment on the mean and the median student Simce scores.

[30]For instance, to analyse the impact on ranking we show the predicted change in the standardised Simce scores ranking when a student i is taught by a teacher j' who is 1 SD more effective than her current teacher j, holding everything else constant.

Table 5.6: Distribution of Teacher Effects, Hypothetical Treatment and Teacher Ranking
4^{th} grade 2005 selected sample cohort

Part A. Distribution of Teacher Effects

	Lambda assigned (λ=0.4) **Model 4**		*Lambda estimated* **Model 1**	
	Stdsed. Language score (1)	Stdsed. Maths score (2)	Stdsed. Language score (3)	Stdsed. Maths score (4)
Mean TEs	0.0	0.0	0.0	0.0
SD TEs	**0.21**	**0.31**	**0.19**	**0.28**
Percentiles (TEs)				
10%	-0.27	-0.39	-0.24	-0.36
25%	-0.14	-0.21	-0.13	-0.19
50%	0.00	0.00	0.00	0.01
75%	0.15	0.21	0.13	0.19
90%	0.27	0.39	0.23	0.36

Part B. Hypothetical treatment: Being exposed to 1 SD more effective teacher

	Movement in percentile Ranking (1)	(2)	Movement in percentile Ranking (3)	(4)
Pupil in the Mean	46th to 55th	48th to 60th	46th to 54th	48th to 59th
Pupil in the Median	50th to 59th	50th to 62nd	50th to 58th	50th to 61th

Part C. Teacher Ranking: Correlation analysis

	Model 4 & Model 1		
	Spearman's rho	p-value	Obs
Individual rank	0.98	0.00	3,151

Notes: **(i)** The pupil in the Mean (and the in Median) is with respect to the Standardised Simce score for each subject. Those values can be observed from Table: "Simce score distribution - 4th Grade 2005 selected sample cohort". **(ii)** The Mean TEs and SD TEs are obtained from the empirical Bayes distribution. **(iii)** Columns (1) and (3) refers to Standardised Language scores for Model 4 and 1, respectively, while columns (2) and (4) correspond to Standardised Maths scores for Model 4 and 1, respectively. **(iv)** General teachers have the same ranking either on Language or Maths teacher effects.

more effective teacher, she would jump up to the 59^{th} and 62^{nd} percentiles in Language and Maths, respectively (when the TEs are estimated under **Model 4**). Similarly, the expected jumps using **Model 1** estimations would be up to the 58^{th} and 61^{nd} percentiles, respectively.

To compare how **Models 4** and **1** differ constructing teacher rankings based on TE estimates, we analyse the correlation between the rankings estimating the *Spearman's rho* coefficient. In Table 5.6 (Part C), we show the teacher ranking correlation is 0.98. This estimated coefficient suggests that, even if Model 1 is inconsistent due to the potential correlation of the lagged dependent variable and the error term ε, it predicts similarly the sorting of teachers based on their TE estimates.[31]

The results obtained from **Models 4** and **1**, where differences in the pre-

[31] Although we do not report the results of teacher effect estimates, it is important to highlight that the estimated Spearman's rho between **Model 3** ($\lambda = 1$) and the other two (**Models 1** and **4**) is around 0.31. This rank correlation shows that the predictions of teacher effects can be affected by the omitted lagged dependent variable and individual student ability from the model.

dicted ranking transitions between subjects, suggest that general teachers are more capable of improving pupil performance in Maths than in Language. This finding is based on the expected jumps in Language (around 8-9 percentile positions) versus the predicted movements in Maths (around 11-12 percentile positions), when the student in the mean (or the median) is taught by a more effective teacher.

We do not explore differences on the predicted impact by pupil's gender, but we propose TEs might differ when teachers teach in a single-sex environment present the estimation results for single-sex school sub-samples. Then, we also study teacher effect heterogeneities by type of school dependence.

In both cases, we want to check whether our estimates are common across different school sub-samples or whether TEs are less or more dispersed in some educational environments than in others. If it emerges that there are not significant differences, then we assure that our models are robust and representative of different type of schooling education. However, if instead we find particular cases where estimated TE distributions significantly differ, that would merit further investigation on explaining why these differences might appear.

Single-sex schools

Differences in academic achievement have been observed between girls and boys, where girls usually outperform boys in humanistic areas, while boys fare better in scientific ones. Particularly, from the estimation results obtained in our VAMs (see Table 5.4), we infer that, on average girls tend to perform worse in Maths than boys. The opposite occurs in Language, although the difference here is smaller. This fact suggests the existence of what it is known as a *gender gap*. Authors have studied whether being in single-sex schools environment or having the same-gender teacher-student would have positive effects on pupil academic performance. However, Doris et al. (2013); Carrington et al. (2008); Winters et al. (2013), found no conclusive evidence of being benefited by these types of treatment.

The *gender gaps* could be explained by differences in a subject's appeal between girls and boys. There might also be differences in the level of effort exerted by teacher when they are teaching girls and boys in specific subjects. Additionally, single-sex classrooms could be less disruptive (especially in only-girls classes), and that might enhance pupils and teacher attention.

We analyse whether there is any advantage in terms of teacher effectiveness when a teacher is teaching in a single-sex school, and hence a single-sex classroom. To check for differences between the initial sample and the single-sex school sub-samples, we estimate TEs and compare the predicted impact on their respective Simce scores distribution.

The sub-samples are relatively small, with considerably fewer teachers and schools observations (specially for boys-only schools). To have an idea of how different the sub-samples are between them, we show the descriptive statistics in Table 5.7.[32]

From Table 5.7, we observe non-important differences in terms of GPA and school marks, but the Maths Simce scores mean is noticeably higher in boys-only schools. The difference between subjects is confirmed when we compare the standardised Maths Simce scores means, being 0.3 for girls-only schools and 0.6 for boys-only schools. With respect to the family background variables, the sub-sample of girls-only schools seems to be disadvantaged compared to boys-only schools, as the average education level reached for both parents and the level of household income is lower in girls-only schools.

Among class level variables, there are no noticeable differences, while at teacher level variables we find remarkable differences such as in the proportion of female teachers, which is 99% in girls-only schools while it is just 79% in boys-only schools, even lower than the 90% observed from the original cohort and the initial selected sample (see Table 5.3). Additionally, the average years of teaching experience in girls-only schools is more than four years higher than teachers in boys-only schools.

The descriptive statistics of school level variables suggest some differences in the type of school dependence between single-sex schools. Girls-only schools are mainly *Municipal* schools, while boys-only schools are principally *Private Voucher* schools. Both sub-samples have little presence in rural areas, with less than 3% of their schools. In terms of number of classes per grade (equivalent to the number of general teachers per grade), both types of school have similar averages. The average of students per grade (or cohort size) is larger in girls-only schools than boys-only.

[32] It is important to have in mind that these sub-samples satisfy all the requirements for the original selection sample discussed previously.

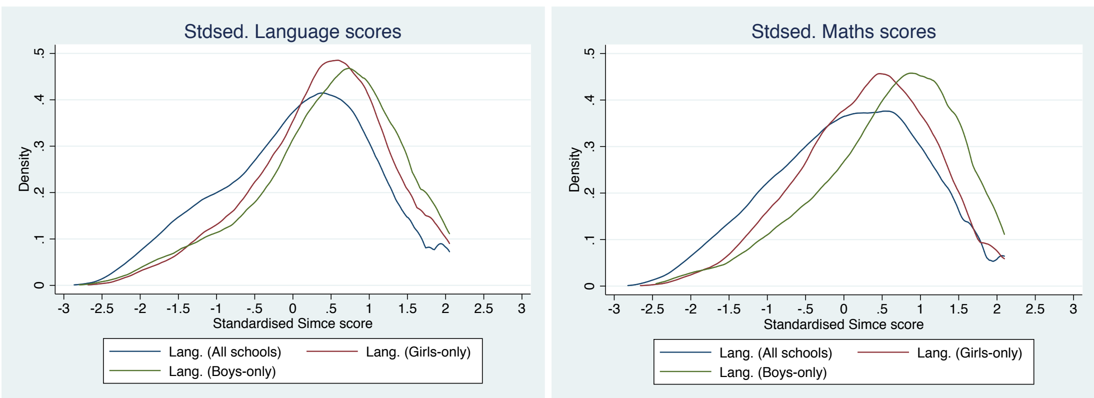
Table 5.7: Descriptive statistics - Single-sex schools
4^{th} grade 2005 selected sample cohort

Girls-only schools	Mean	Std. Dev.	Min	Max
Pupil Level				
GPA 4th 2005	6.0	0.5	1	7
GPA 3th 2004	6.0	0.7	1	7
School Language Marks 2005	5.6	0.7	2.9	7
School Maths Marks 2005	5.4	0.8	2.7	7
School Language Marks 2004	5.8	0.7	2.55	7
School Maths Marks 2004	5.6	0.8	2.4	7
Language Simce Scores 4th 2005	275.0	46.9	112.8	364.8
Maths Simce Scores 4th 2005	264.1	48.1	100.8	363.6
Stdsed. Lang. Simce Score 4th 2005	0.4	0.9	-2.7	2.1
Stdsed. Maths Simce Score 4th 2005	0.3	0.9	-2.7	2.1
Gender (Female=1)	1.0	0.0	1.0	1.0
Age	9.1	0.3	8	12
Attendance	94.4	6.0	0	100
Speecial Needs	0.0	0.0	0	1
Mother Education	3.4	1.9	0	8
Father Education	3.6	2.1	0	8
Household Income	2.6	2.8	0	12
Class Level				
Class size	31.3	7.0	15	45
Peers Average GPA	5.9	0.3	5.089	6.603
Teacher Level				
Gender (Female=1)	0.99	0.1	0	1
Years of experience in the system	23.5	10.3	1	40
(Teaching hrs / Contract hrs) Ratio	0.9	0.1	0.38	1
School Level				
Municipal Schools	0.5	0.5	0	1
Private Voucher Schools	0.4	0.5	0	1
Unsubsidised Private Schools	0.1	0.3	0	1
Rural Area	0.0	0.2	0	1
Number of classes per grade	2.5	0.7	2	5
Number of students per grade	76.8	26.1	31	154
Number of students	5,070			
Number of teachers	162			
Number of schools	66			

Boys-only schools	Mean	Std. Dev.	Min	Max
Pupil Level				
GPA 4th 2005	5.9	0.6	1	7
GPA 3th 2004	6.0	0.7	1	7
School Language Marks 2005	5.6	0.8	2.6	7
School Maths Marks 2005	5.5	0.8	2.7	7
School Language Marks 2004	5.8	0.7	2.1	7
School Maths Marks 2004	5.7	0.8	2.15	7
Language Simce Scores 4th 2005	279.9	49.6	106.7	364.8
Maths Simce Scores 4th 2005	279.5	50.8	112.0	363.6
Stdsed. Lang. Simce Score 4th 2005	0.5	0.9	-2.8	2.1
Stdsed. Maths Simce Score 4th 2005	0.6	0.9	-2.5	2.1
Gender (Female=1)	0.0	0.0	0.0	0.0
Age	9.2	0.4	9	12
Attendance	94.8	4.7	0	100
Speecial Needs	0.0	0.0	0	1
Mother Education	4.0	2.1	0	8
Father Education	4.2	2.3	0	8
Household Income	4.0	3.7	0	12
Class Level				
Class size	29.6	7.5	15	45
Peers Average GPA	5.9	0.3	5.059	6.517
Teacher Level				
Gender (Female=1)	0.79	0.4	0	1
Years of experience in the system	18.8	12.4	0	40
(Teaching hrs / Contract hrs) Ratio	0.8	0.2	0.29	1
School Level				
Municipal Schools	0.3	0.5	0	1
Private Voucher Schools	0.5	0.5	0	1
Unsubsidised Private Schools	0.2	0.4	0	1
Rural Area	0.0	0.2	0	1
Number of classes per grade	2.4	0.5	2	4
Number of students per grade	70.8	22.2	31	120
Number of students	2,547			
Number of teachers	86			
Number of schools	36			

Notes: (i) The selection sample was made for identification, and the most important sources of dropped observations were schools with less than two classes per grade and schools and schools with specialised teachers only, both sum up to 20% of the original pupil's observations in the cohort. **(ii)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(iii)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600.

Figure 5.5: Kernel distributions: Stdsed. Simce Scores - Single-sex schools
4^{th} grade 2005 selected sample cohort



Graphically, from Figure 5.5, we confirm the differences in Simce scores between single-sex schools. Only-boys schools present a larger concentration of higher scores in Language and Maths, although the difference is considerably larger in Maths.

Acknowledging the observable differences between both sub-samples, we analyse the distribution of Simce scores to compare the impact of TEs on their respective distributions.

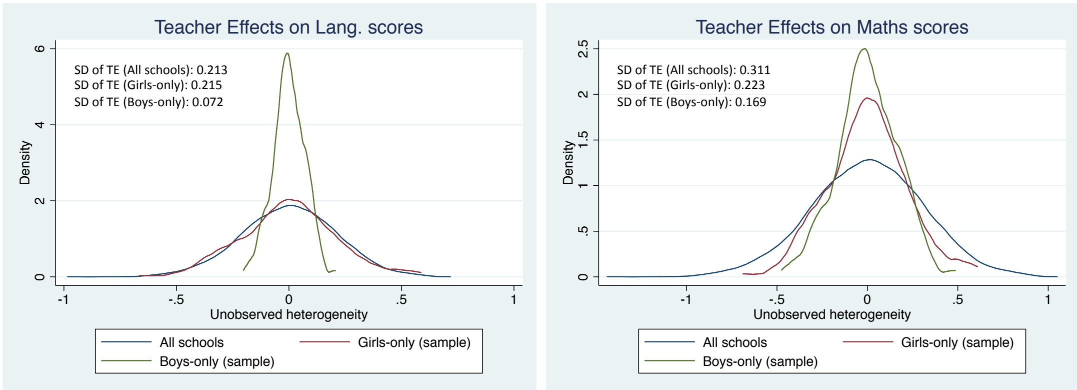
Table 5.8: Simce score distribution by gender-type of school
4^{th} grade 2005 selected sample cohort

		All Schools		Girls-only schools		Boys-only schools	
Percentile		Language	Maths	Language	Maths	Language	Maths
10th	*Absolute Score*	183.4	176.3	209.4	198.6	208.9	207.1
	Standardised Score	-1.36	-1.29	-0.87	-0.89	-0.88	-0.73
25th	*Absolute Score*	222.8	213.0	246.5	231.9	251.7	248.2
	Standardised Score	-0.62	-0.63	-0.17	-0.29	-0.07	0.01
50th	*Absolute Score*	263.8	254.2	279.7	267.7	286.6	286.6
	Standardised Score	0.15	0.12	0.45	0.36	0.58	0.70
75th	*Absolute Score*	296.5	291.3	307.6	299.1	315.9	317.6
	Standardised Score	0.77	0.79	0.98	0.93	1.13	1.26
90th	*Absolute Score*	323.5	320.1	333.2	324.2	339.6	341.3
	Standardised Score	1.28	1.31	1.46	1.38	1.58	1.69
Abs. Mean		258.5	251.1	275.0	264.1	279.9	279.5
Abs. SD		**52.3**	**54.2**	**46.9**	**48.1**	**49.6**	**50.8**
Stzd. Mean		0.054	0.062	0.364	0.296	0.456	0.574
Stdz. SD		**0.983**	**0.980**	**0.881**	**0.871**	**0.932**	**0.918**
Number of Schools			1,337		66		36
Number of Teachers			3,151		162		86
Number of Students			89,710		5,072		2,547

In Table 5.8, we recognise that given the absolute (and standardised) Simce scores, students from single-sex schools clearly show better performance than the initial selected sample (all schools). Nevertheless, we want to check whether teacher effectiveness varies between these types of schools, and to analyse the results we focus on estimations made with **Model 4**, using a preset persistence parameter λ_0 fixed at 0.4. The table of results with the MLE estimations for both sub-samples and all model specifications are presented in Appendix 5.2, and the tables with the estimated EB distributions for all unobserved heterogeneities are presented in Appendix 5.4.

Regarding the estimated TEs, we show in Figure 5.6 how different the distributions of TEs between girls-only and boys-only schools are, particularly, on Language scores. TEs in boys-only schools seem to be more concentrated around the mean 0, with a considerably lower SD of estimated TEs on both Language and Maths scores. While, the distribution of TEs on Language for girls-only schools is very similar to the distribution observed for the TEs in all schools, both with a SD of 0.21. The distributions of TEs is closer between single-sex of schools in Maths scores, although TEs in boys-only schools are still more concentrated around the mean.

Figure 5.6: Empirical Bayes distribution of Teacher Effects by gender-type of school
4^{th} grade 2005 selected sample cohort



Note: School groups in brackets refer to the corresponding school sub-sample.

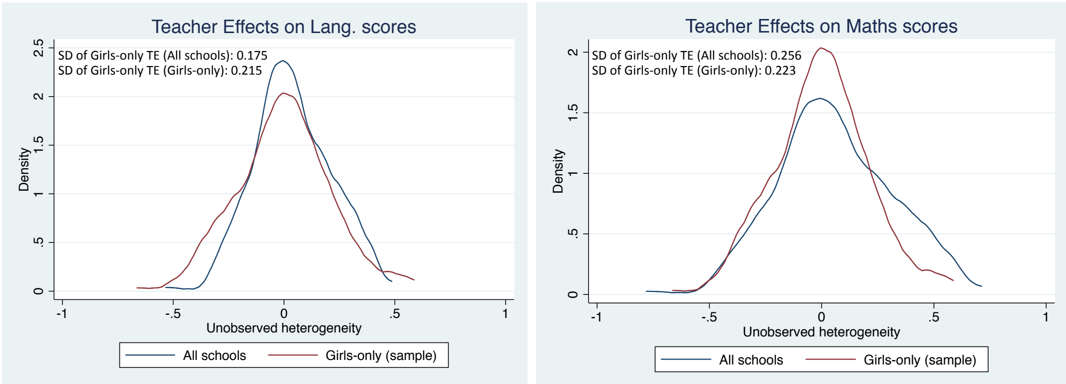
From both sub-samples, we observe lower dispersion of TEs on Maths compared to the estimation obtained with all schools, which suggests teachers might be more homogeneous in single-sex schools. However, the striking difference is found in the distribution of TEs on Language scores, as the dispersion of the TEs in boys-only schools is remarkably shrunk towards the mean. Explanations can be addressed from the teacher or pupil's perspective. Either teachers recruited in boys-only schools are more similar in teaching skills on Language (and Maths), or potential student selection on this type of schools has implied a concentration of more homogenous, and high performers, group of pupils. In any of these cases, it might be difficult to find large differences on pupil's achievement when changing one teacher to more effective within this sub-sample.

In terms of observable characteristics, we found, in Table **??** some considerable differences between girls-only and boys-only schools, particularly on academic performance where boys outperform girls. However, with respect to the observable teacher characteristics we find teachers in boys-only schools seem to be even more heterogeneous than teachers girls-only schools, and if they are more heterogeneous it might be due to unobserved characteristics.

We are aware that estimations of TEs might be affected by sample selection, and their samples might be not representative at all given the few schools and teachers considered in the estimation, especially in boys-only schools. Therefore, we focus exclusively on heterogeneous teacher effects on girls-only schools and we check whether the estimation of TE estimates differs in this type of schools when we use different samples, the selected sub-sample and the original one. In Figure 5.7, we observe that using the whole sample of schools, the estimated TEs on

Language are more concentrated around the mean, while the predicted TEs on Maths are more concentrated in the right-hand side of the distribution.

Figure 5.7: Empirical Bayes distribution of Teacher Effects within girls-only schools
4^{th} grade 2005 selected sample cohort



Note: School groups in brackets refer to the corresponding school sub-sample.

Note: School groups in brackets refer to the corresponding school sub-sample.

However, to check whether a single-sex school environment improves teacher effectiveness, we compare the impact of TEs in girls-only schools with the whole sample in terms of expected pupil's ranking movements. We analyse the results shown in Table 5.9 (Part B), where we present the predicted jumps of a pupil in the mean (and in the median) of the Simce scores distribution, when she is exposed to the hypothetical treatment of being taught by a more effective teacher.

Girls-only classrooms seem to boost the teacher effectiveness impact on Language scores. A pupil in the mean or in the median, exposed to 1 SD more effective teacher, improve her percentile ranking in two positions more than the mean or median pupil from all schools sample. The expected jumps on Language scores ranking within all schools are from 46^{th} to 55^{th} for the mean pupil, and from 50^{th} to 59^{th} for the median one, while the predicted jumps in the girls-only schools are up to 57^{th} and the 61^{st}, respectively. On the contrary, the predicted movements on Maths score ranking are lower for girls-only schools (also for boys-only schools) in comparison to the expected impact with the whole sample.

We also analyse whether the predicted ranking among teachers from girls-only schools differ considerably when the estimation is made with the sub-sample compared to the initial selected sample. In Table 5.9 (Part C), we test the correlation of both rankings estimating the *Spearman's rho* coefficient. This correlation ranking consists in comparing how similar teacher are sorted based on their estimated TEs when two different samples are employed in the estimation. In this particular case, even if the estimated rankings are significantly positively correlated, the 0.60 of the ranking correlation between both estimation samples, suggest considerable differences in the teachers' rank prediction based on estimated TEs.

Table 5.9: Distribution of Teacher Effects, Hypothetical Treatment and Teacher Ranking
4^{th} grade 2005 selected sample cohort by gender-type of schools

Part A. Distribution of Teacher Effects

	All Schools **Model 4** *(λ=0.4)*		*Girls-only schools* **Model 4** *(λ=0.4)*		*Boys-only schools* **Model 4** *(λ=0.4)*	
	Stdsed. Lang. score (1)	Stdsed. Maths score (2)	Stdsed. Lang. score (3)	Stdsed. Maths score (4)	Stdsed. Lang. score (5)	Stdsed. Maths score (6)
Mean TEs	0.0	0.0	0.0	0.0	0.0	0.0
SD TEs	**0.21**	**0.31**	**0.21**	**0.22**	**0.07**	**0.17**
Percentiles (TEs)						
10%	-0.27	-0.39	-0.13	-0.13	-0.04	-0.09
25%	-0.14	-0.21	0.01	0.01	0.00	0.00
50%	0.00	0.00	0.13	0.13	0.05	0.12
75%	0.15	0.21	0.27	0.28	0.09	0.22
90%	0.27	0.39	0.37	0.38	0.11	0.27

Part B. Hypothetical treatment: Being exposed to 1 SD more effective teacher

	Movement in percentile Ranking (1)	(2)	Movement in percentile Ranking (3)	(4)	Movement in percentile Ranking (5)	(6)
Pupil in the Mean	46th to 55th	48th to 60th	46th to 57th	48th to 58th	45th to 48th	44th to 52nd
Pupil in the Median	50th to 59th	50th to 62nd	50th to 61st	50th to 61st	50th to 54th	50th to 58th

Part C. Teacher Ranking: Correlation analysis

	All schools & Single-sex schools		
	Spearman's rho	p-value	Obs
Only-girls teacher rank	0.60	0.00	162
Only-boys teacher rank	0.82	0.00	82

Notes: (i) The pupil in the Mean (and the in Median) is with respect to the Standardised Simce score for each subject. Those values can be observed from Table: "Simce score distribution - 4^{th} Grade 2005 selected sample cohort". **(ii)** The Mean TEs and SD TEs are obtained from the empirical Bayes distribution. **(iii)** Both sub-samples were estimated under Model 4, where columns (1), (3) and (5) refer to Standardised Language scores for Model 4, and columns (2), (4) and (6) correspond to Standardised Maths scores. **(iv)** General teachers have the same ranking either on Language or Maths teacher effects.

Although we find there might be an increase in the teacher's impact on Language score in girls-only schools, it is important to emphasise that girls-only schools represent 6% of the whole sample (and boys-only schools just 3%). Moreover, the prediction of teachers ranking using the only-girls schools sub-sample differ from the ranking obtained with the whole sample.

Therefore, the results we present here for single-sex schools might not be so robust given the small size of the sub-samples, and we should be cautious with any inferences made based on these particular findings.

In the following subsection, we analyse potential differences of TEs predictions, given different types of schools sub-samples.

Heterogeneous Teacher Effects by types of school

To estimate TEs by MLE-EB, we have assumed random assignment of students and teachers to schools. We also introduce a rich set of student, teacher and school level variables to control for potential sources of observable non-random assignment. In addition, we rely on the non-random assignment of students to classrooms and teachers to schools to have a consistent estimator for **Model 4**.

However, it is important to consider that, in Chile, schools are classified by their type of dependence in terms of funding and managerial control. *Munici-*

pal and *Private Voucher* schools receive public funds, while *Private Unsubsidised* schools are exclusively privately funded. In the Introductory Chapter, we mentioned that *Municipal* schools are not allowed to select pupils, and their teacher contracts are ruled by the National Teacher Statute (NTS). These facts suggest that random assignment assumption are more difficult to hold in private schools.

Thus, we check whether estimation of TEs from *Municipal* schools differ significantly from the original sample presented earlier. We also estimate TEs in a sub-sample which removes only *Private Unsubsidised* schools, as *Municipal* and *Private Voucher* schools are oriented to serve a similar population target.

We do not estimate TEs for *Private Unsubsidised* schools only because the size of the sub-sample is very small. Similar to the case of boys-only schools, we could not make fully reliable inferences on estimates from this group.

When we describe only *Municipal* schools, and *Municipal* joined with *Private Voucher* school sub-samples in Table 5.10, we do not find significant differences in pupils academic performance, although the averages in Simce scores are a little higher when including *Private Voucher* schools. Proportion of female pupils and the rates of attendance are basically the same between both sub-samples.

Graphically, we can also observe from Figure 5.8 that **Municipal** schools accumulate more students below the mean of the standardised Simce scores. Hence, the distribution of students from the sub-sample that includes *Private Voucher* schools is shifted to the right in both subjects. *Municipal* schools concentrate more a low-achievement pupils in Language and Maths, although the shape of the distributions is very similar within subjects.

In terms of parental education and household income, *Municipal* schools show marginally lower averages. Class level variables are for both subsamples very similar, from teacher observable variables *Municipal* schools seem to have an average 4 years of more experience. In terms of school size, given by the number of classes per grade and students per grade, there are non-remarkable differences either.

Apart from the shifted Simce scores distributions and the teachers' years of experience, both groups of schools seem to be very similar. Then, the comparison of TE estimates between both sub-samples would allow to infer whether the estimation of TE's impact on pupil achievement differ when private schools are considered in the estimation.

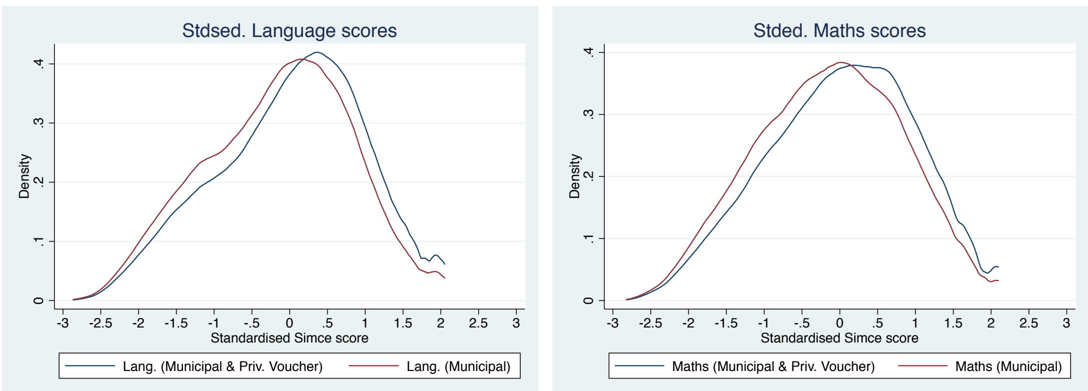
Table 5.10: Descriptive statistics by type of school dependence
4^{th} grade 2005 selected sample cohort

Munic. and Priv. Voucher schools	Mean	Std. Dev.	Min	Max
Pupil Level				
GPA 4th 2005	5.9	0.6	1	7
GPA 3th 2004	5.9	0.8	1	7
School Language Marks 2005	5.5	0.8	1	7
School Maths Marks 2005	5.4	0.9	1	7
School Language Marks 2004	5.7	0.8	1	7
School Maths Marks 2004	5.5	0.9	1	7
Language Simce Scores 4th 2005	256.4	51.8	103.1	364.8
Maths Simce Scores 4th 2005	248.8	53.6	91.7	363.6
Stdsed. Lang. Simce Score 4th 2005	0.0	1.0	-2.9	2.1
Stdsed. Maths Simce Score 4th 2005	0.0	1.0	-2.8	2.1
Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.1	0.4	8	12
Attendance	94.4	5.8	0	100
Speecial Needs	0.0	0.1	0	1
Mother Education	2.7	1.9	0	8
Father Education	3.0	2.1	0	8
Household Income	1.8	1.9	0	12
Class Level				
Class size	28.8	6.7	15	45
Peers Average GPA	5.7	0.4	2.4	6.7
Teacher Level				
Gender (Female=1)	0.89	0.3	0	1
Years of experience in the system	20.3	11.5	0	40
(Teaching hrs / Contract hrs) Ratio	0.90	0.1	0.11	1
School Level				
Municipal Schools	0.6	0.5	0	1
Private Voucher Schools	0.4	0.5	0	1
Unsubsidised Private Schools	0.0	0.0	0	0
Rural Area	0.0	0.2	0	1
Number of classes per grade	2.3	0.6	2	5
Number of students per grade	67.5	25.2	30	209
Number of students	85,737			
Number of teachers	2,974			
Number of schools	1,271			

Municipal schools	Mean	Std. Dev.	Min	Max
Pupil Level				
GPA 4th 2005	5.8	0.6	1	7
GPA 3th 2004	5.9	0.8	1	7
School Language Marks 2005	5.4	0.8	1	7
School Maths Marks 2005	5.3	0.9	1	7
School Language Marks 2004	5.6	0.8	1	7
School Maths Marks 2004	5.5	0.9	1	7
Language Simce Scores 4th 2005	247.5	51.2	103.1	364.8
Maths Simce Scores 4th 2005	239.8	53.2	91.7	363.6
Stdsed. Lang. Simce Score 4th 2005	-0.2	1.0	-2.9	2.1
Stdsed. Maths Simce Score 4th 2005	-0.1	1.0	-2.8	2.1
Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.2	0.4	8	12
Attendance	94.2	5.9	0	100
Speecial Needs	0.0	0.1	0	1
Mother Education	2.3	1.8	0	8
Father Education	2.6	2.1	0	8
Household Income	1.3	1.5	0	12
Class Level				
Class size	27.4	6.3	15	45
Peers Average GPA	5.6	0.3	2.3	6.7
Teacher Level				
Gender (Female=1)	0.91	0.3	0	1
Years of experience in the system	24.2	10.3	0	40
(Teaching hrs / Contract hrs) Ratio	0.90	0.1	0.11	1
School Level				
Municipal Schools	1.0	0.0	1	1
Private Voucher Schools	0.0	0.0	0	0
Unsubsidised Private Schools	0.0	0.0	0	0
Rural Area	0.0	0.2	0	1
Number of classes per grade	2.4	0.6	2	5
Number of students per grade	64.6	23.2	30	181
Number of students	51,270			
Number of teachers	1,872			
Number of schools	794			

Notes: **(i)** The selection sample was made for identification, and the most important sources of dropped observations were schools with less than two classes per grade and schools and schools with specialised teachers only, both sum up to 20% of the original pupil's observations in the cohort. **(ii)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(iii)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600.

Figure 5.8: Kernel distributions: Stdsed. Simce Scores by type of school dependence
4^{th} grade 2005 selected sample cohort



The predicted impact of TEs depends on the standardised Simce score distribution from which the estimation was made, conditional on the standardised Simce scores mean. In Table 5.11, we describe the summary distribution of Simce scores for both school sub-samples compared to the initial selected cohort.

Table 5.11: Simce score distribution by type of school dependence
4^{th} grade 2005 selected sample cohort

		All Schools		Municipal and Priv. Voucher		Only Municipal	
Percentile		Language	Maths	Language	Maths	Language	Maths
10th	*Absolute Score*	183.4	176.3	182.1	174.9	175.5	167.5
	Standardised Score	-1.36	-1.29	-1.38	-1.32	-1.51	-1.45
25th	*Absolute Score*	222.8	213.0	220.8	211.0	210.6	201.1
	Standardised Score	-0.62	-0.63	-0.65	-0.66	-0.85	-0.84
50th	*Absolute Score*	263.8	254.2	261.5	251.7	251.6	241.3
	Standardised Score	0.15	0.12	0.11	0.07	-0.07	-0.12
75th	*Absolute Score*	296.5	291.3	294.0	288.5	284.8	279.1
	Standardised Score	0.77	0.79	0.72	0.74	0.55	0.57
90th	*Absolute Score*	323.5	320.1	320.4	317.3	311.1	309.0
	Standardised Score	1.28	1.31	1.22	1.26	1.04	1.11
Abs. Mean		258.5	251.1	256.4	248.8	247.5	239.8
Abs. SD		**52.3**	**54.2**	**51.8**	**53.6**	**51.2**	**53.2**
Stzd. Mean		0.055	0.062	0.014	0.019	-0.152	-0.143
Stdz. SD		**0.983**	**0.980**	**0.974**	**0.969**	**0.963**	**0.962**
Number of Schools			1,337		1,271		794
Number of Teachers			3,151		2,974		1,872
Number of Students			89,710		85,754		51,283

Average Simce scores in *Municipal* schools are the lowest among all the school sub-groups. The sample of *Municipal* and *Private Voucher* schools has basically the same Simce scores distribution as the original selected cohort, and it might be due to the small size of *Unsubsidised Private* schools within this sample.

To verify whether TE estimates differ when considering private schools compared with estimations based on *Municipal* schools only, we analyse the following aspects: (i) the distribution of TEs in both sub-samples; (ii) the expected transitions in the percentile ranking from the mean and median student; (iii) the teacher ranking correlation between sub-samples. To carry out these analyses, we focus on the TEs estimation using **Model 4**, where λ_0 is preset to a fixed value. The coefficient estimates and the rest of the MLE estimation results are shown in Appendix 5.2.

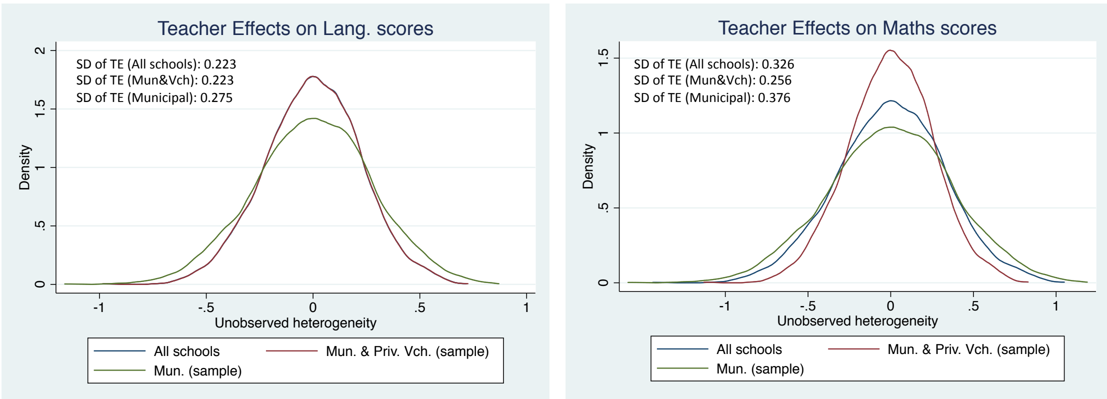
First, we show in Figure 5.9 the empirical Bayes distribution of TEs for both school groups.[33]

In both subjects, there is a larger dispersion of TEs in *Municipal* schools, compared to all schools and the other sub-sample. This fact suggests that teachers within *Municipal* schools are more heterogeneous than in private schools, and it

[33]The tables with the EB distribution for all latent variables and all estimation models are presented in Appendix 5.3.

is more likely to find low and high effective teachers within this group. The distributions of TEs with and without *Unsubsidised Private* schools only differ on Maths scores, where TEs estimations are more dispersed considering the whole school sample.

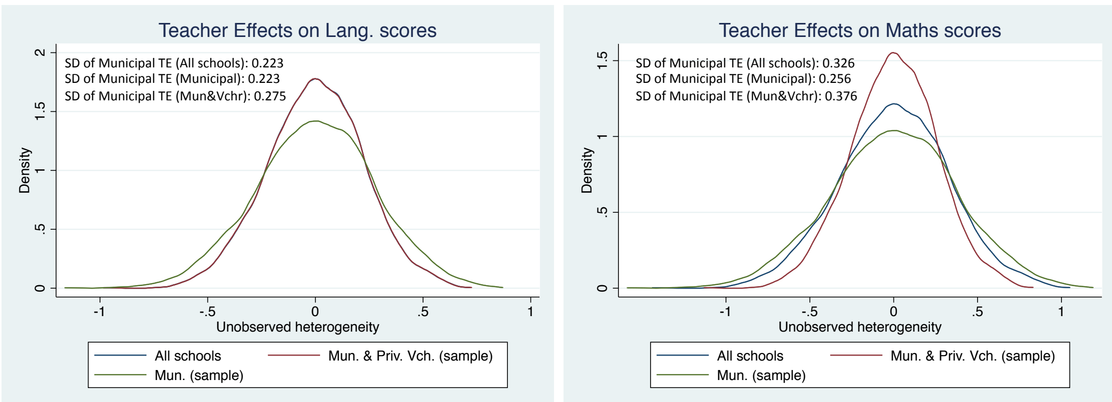
Figure 5.9: Empirical Bayes distribution of Teacher Effects by type of school dependence
4^{th} grade 2005 selected sample cohort



Note: School groups in brackets refer to the corresponding school sub-sample, where "Mun&Vch" refers to the sub-sample of *Municipal* and *Private Voucher* schools.

Focusing only on TEs of *Municipal* school teachers, we compared the EB distributions within this group in Table 5.10. In terms of dispersion and shape, the distributions of Municipal TEs, on Language and Maths, do not differ from the estimations of TEs using the whole sample, and without including the *Unsubsidised Private* schools. This fact suggest that estimations of TEs are similar, or non-noticeably different, when we consider private schools in the estimation sample, which are supposed to be more likely to violate random assignment assumptions.

Figure 5.10: Empirical Bayes distribution of Teacher Effects within Municipal schools
4^{th} grade 2005 selected sample cohort



Note: School groups in brackets refer to the corresponding school sub-sample, where "Mun&Vch" refers to the sub-sample of *Municipal* and *Private Voucher* schools.

From Table 5.12, Part A, we observe no differences in terms of SD of TEs between all schools and the estimates from the sample without *Unsubsidised Private* schools. However, from Figure 5.10, we are aware that the shape of the distribution varies between them, what as confirmed in the percentile table of Part A. This is also the case for TE estimates in *Municipal* schools only, where the Simce score distribution and the EB distribution of TEs differ from the rest of the groups. Therefore, we analyse whether the predicted impact on pupils ranking change among the samples.

Table 5.12: Distribution of Teacher Effects, Hypothetical Treatment and Teacher Ranking
4^{th} grade 2005 selected sample cohort by type of school dependence

	All Schools **Model 4** *(λ=0.4)*		*Municipal and Priv. Voucher schools* **Model 4** *(λ=0.5)*		*Municipal schools* **Model 4** *(λ=0.5)*	
	Stdsed. Lang. score (1)	Stdsed. Maths score (2)	Stdsed. Lang. score (3)	Stdsed. Maths score (4)	Stdsed. Lang. score (5)	Stdsed. Maths score (6)
Mean TEs	0.0	0.0	0.0	0.0	0.0	0.0
SD TEs	**0.21**	**0.31**	**0.22**	**0.31**	**0.28**	**0.38**
Percentiles (TEs)						
10%	-0.27	-0.39	-0.27	-0.37	-0.36	-0.49
25%	-0.14	-0.21	-0.14	-0.19	-0.18	-0.25
50%	0.00	0.00	0.00	0.01	0.00	0.00
75%	0.15	0.21	0.15	0.20	0.19	0.26
90%	0.27	0.39	0.27	0.36	0.35	0.47

Part B. Hypothetical treatment: Being exposed to 1 SD more effective teacher

	Movement in percentile Ranking (1)	(2)	Movement in percentile Ranking (3)	(4)	Movement in percentile Ranking (5)	(6)
Pupil in the Mean	46th to 55th	48th to 60th	46th to 55th	48th to 60th	47th to 58th	49th to 64th
Pupil in the Median	50th to 59th	50th to 62nd	50th to 59th	50th to 62nd	50th to 61st	50th to 65th

Part C. Teacher Ranking: Correlation analysis

	All schools & Mun. and Priv. Voucher			All schools & Municipals		
	Spearman's rho	p-value	Obs	Spearman's Rho	p-value	Obs
Private Voucher teacher rank	1.00	0.00	1,102			
Muncipal teacher rank	1.00	0.00	1,872	0.99	0.00	1,872

Notes: **(i)** The pupil in the Mean (and the in Median) is with respect to the Standardised Simce score for each subject. Those values can be observed from Table: "Simce score distribution - 4^{th} Grade 2005 selected sample cohort". **(ii)** The Mean TEs and SD TEs are obtained from the empirical Bayes distribution. **(iii)** Both sub-samples were estimated under Model 4, where columns (1), (3) and (5) refer to Standardised Language scores for Model 4, and columns (2), (4) and (6) correspond to Standardised Maths scores. **(iv)** General teachers have the same ranking either on Language or Maths teacher effects.

In Part B of Table 5.12, we show the expected changes in the percentile ranking when a pupil from the mean and the median is exposed to a more effective teacher. We find no differences in predicted pupil movements between the samples either considering or not considering *Unsubsidised Private* schools. Nevertheless, the impact of TEs on the percentile ranking within *Municipal* schools is strictly larger on Language and Maths, either for the pupil in the mean or in the median. The expected movement in the Language score ranking corresponds to 11 percentile positions, two more than in the whole sample. While, the predicted jump on Maths ranking is up to 15 percentile positions, larger than the increase of 12 positions predicted for the original sample.

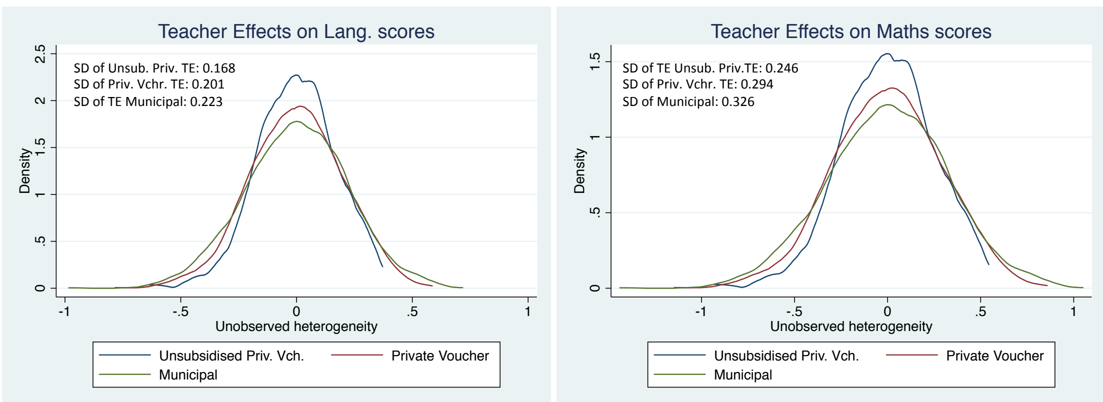
The fact that TEs are more heterogenous in *Municipal* schools, imply they have larger impact on pupil achievement. Hence, in this type of schools, which concentratea more on low achievement and less advantaged pupils, the effect of teachers it might be even more relevant than in private schools.

Finally, to check whether TE estimations differ in terms of teacher ranking predictions when we include *Private Voucher* schools in the sample, we test for the ranking correlation of teachers in *Municipal* schools under different sub-samples. In Table 5.12 (Part C), we compare the teachers ranking between two different pairs of samples: (i) All schools (sample 1) & *Municipal* and *Private Voucher* schools (sub-sample 1); (ii) All schools (sample 1) & *Municipal* schools only (sub-sample 2).

For the first pair of samples, we estimate the *Spearman's rho* coefficient for the teachers ranking correlation in *Private Voucher* schools and *Municipal* schools independently. We find an almost perfect positive rank correlation between the estimation under this group of samples. In the second pair, the results are similar, the sorting of teacher from *Municipal* schools is basically the same when the are estimated with the whole sample (sample 1 or under *Municipal* schools only (sub-sample 2).

Based on the fact that predicted TE's impact of *Municipal* teachers are equally estimated independently from the sample, we show in Figure 5.11 the TE distributions by type of school dependence obtained from the original selected cohort. Here, we confirm that among type of schools, *Municipal* schools accumulate more teachers in the tails of the EB distribution of TEs, with more high effective but also more low effective teachers.

Figure 5.11: Empirical Bayes distribution of Teacher Effects separated by type of school
4^{th} grade 2005 selected sample cohort



Note: "Priv. Vchr." refers to the *Private Voucher* schools; "Unsub. Priv." refers to *Unsubsidised Private* schools.

These results suggest that heterogeneous teacher effects can be estimated equally from each sub-sample, at least in our particular context. These findings also support the assumption of non-random assignment of student to teacher in the Chilean school system, as the estimations of TE's impact do not change dramatically under the different types of school samples, and neither do the prediction of teacher rankings.

5.5 Conclusion

In this chapter, we have presented the methodology used to estimate teacher effects from our VAMs. Relying on a particular set of assumptions to consistently estimate the models, we also described the sample selection criteria required to identify student, teacher and school unobserved factors simultaneously.

We constrain our sample to general teachers only from primary schools, and we adopt the MLE-EB approach to estimate TEs for **Models 1**, **3** and **4** which were derived from a general achievement function in Chapter 2. We obtain predictions of unobserved heterogeneities from the EB distributions in each model, focusing in particular on the TEs distribution.

We presented TE estimates in terms of SD of EB distributions, and we discussed their predicted impact on pupil achievement by measuring the expected transition in pupils' ranking of being exposed to a more highly effective teacher. The interpretation of estimated SD of TEs is conditioned on the Simce scores distribution. Thus, we predict the impact on a pupil from the mean or the median of the Simce distribution, when she is taught by a teacher who is one standard deviation more effective. However, we suggest caution in using predictions in expected jumps for measuring changes in teachers Value-Added benefit, and particularly, for evaluating policies regarding improvements in teacher effectiveness.

Based on the results obtained when λ_0 equals 0.4 (**Model 4**) and when λ is estimated in the model (**Model 1**), we find that 4^{th} grade general teachers have larger impacts on pupils Maths achievement than on Language achievement. From Table 5.6, we observe how a pupil, either from the mean or the median, moves up more positions in the Maths percentile ranking than in the equivalent Language ranking, when she is exposed to a teacher who is one standard deviation more effective. If a representative student from the 50^{th} percentile is taught by a teacher who is one standard deviation more effective, the ranking position of that student is expected to increase by between 8 and 9 percentiles in Language, and between 11 and 12 in Maths.

We also estimated TEs for single-sex schools, finding slightly larger impacts on Language performance in pupils from girls-only schools, but lower expected improvements on Maths. We do not observe any improvement on students' ranking from being exposed to a more effective teacher in boys-only schools, compared to the predicted impacts obtained from the whole sample. However, we suggest that the small size of both single-sex schools sub-samples may affect the robustness of these results. The girls-only schools sub-sample correspond to the 6% of the original selected cohort, while the boys-only schools sub-sample accounts for just 3%. Besides, the ranking correlation of teachers in *Municipal schools* is not very high, given the estimated *Spearmans rho* coefficient (0.60). So, any inferences that can be made from these results might be misleading.

To check for heterogeneous teacher effect estimates, we disaggregate the original selected cohort into two subs-samples by type of school dependence. In the first, we remove Unsubsidised Private schools from the sample; in the second, we only consider *Municipal schools*. We find that TEs are more heterogeneous in *Municipal schools*; this implies that municipal teachers have a larger impact (positive or negative) on pupil academic performance.

Hence, in this type of school, which concentrates more on students with histories of low achievement, and who come from less-advantaged settings, the effect of teachers is even more important. We also find that heterogeneous teacher effects can be equally estimated from different types of school samples.

In addition, this finding supports the assumption of non-random assignment of students to teachers in the Chilean school system, as the estimation of *Municipal* TEs impact and teacher rankings were basically the same with and without considering private schools.

The results presented here, supported by the evidence of random assignment of students to teachers shown in Chapter 4, confirm the importance of the potential contribution of teachers on pupil academic performance. Moreover, our Value-Added measures provide a complementary instrument to evaluate teacher effectiveness in the Chilean education system.

As our VAMs allow us to recover not only TE estimates, but also SEs, which might be used as an indicator of school quality, we can also further analyse the SE estimates. In discussing the validity of school quality measures, the literature has particularly criticised the construction of school rankings or school league tables. One of the main concerns regarding the publication of school rankings is the potential misleading inferences that can be made based on specific school positions in the ranking. These run a high risk of misclassifying schools, especially when schools are close to each other in the ranking, due to the standard error of SEs estimates.

Both TEs and SEs may vary across time; several factors can alter them such as exogenous shocks, learning process improvements, institutional changes, and the possibility that different students cohorts make the differences in the estimates. Therefore, it is interesting to check how stable TEs and SEs are for a specific period in the Chilean context.

Analysing the evolution of TE and SE estimates, we would minimise the risk of misclassification in a specific years. Additionally, taking advantage of the availability of the data for an extended period of time will allow us to study teacher and school quality movements instead. However, due to the fact that we cannot estimate TE measures for all teachers every year, in Chapter 6 we will be focused on SE estimates and we will relate them to the types of teacher, in terms of quality and verified specialisation.

Further investigations could be carried out to test differences in the impact of general teachers by type of teacher specialisation. We could investigate whether there are significant differences in TEs across types of teacher specialisation by comparing the effects of teachers teaching the same class-cohort for at least two years, relative to those who are consistently assigned to one specific grade, such as to 4^{th} grade.

Chapter 6

School effects: What determines effectiveness in the long run?

Abstract

In this chapter we present estimates of "School Effectiveness" (SEs) in the context of the literature on school league tables (LTs). We aim to analyse school effects in the long run in order to address uncertainty issues associated with yearly school rankings constructed with single measures of SEs.

We are particularly interested in the evolution of school effectiveness and the factors that explain positive or negative trajectories over time. Our results suggest that neither downward nor upward effectiveness trajectories are explained by observable student, teacher and school characteristics. What seems to drive school high (or low) effectiveness in the long run is the proportion of *High* (or *Low*) quality teachers based on TE estimates.

Contents

6.1 Introduction

It is a commonly held belief that schooling, and school quality in particular, are the key factors affecting individuals' subsequent labour market success. Better schools and higher quality education in general are typically expected to improve human capital. It is also often thought that consequent increments in productivity are likely to lead to lower levels of inequality and potentially higher economic growth (Coleman et al. (1966)). Thus, school quality is of significant private and public interest, and drives parental decisions, and motivates key educational policies and reforms.

Earlier investigations in education topics have set schools as one of the most important factors for the focus of education research. However, schools form part of a very complex institutional system which includes pupils, families and teachers within a specific educational environment. This educational context might be influenced by principals and other public regulations. Thus, it has been a great challenge for all disciplines to address properly school quality analysis.

During recent years, the role and importance of teachers have also attracted major attention in many educational studies, with research on teachers' contributions to pupils' academic outcomes. Although the identification of these "Teacher Effects" (TEs) is still widely debated. As discussed the earlier Chapters of this thesis, the Value-Added concept is commonly accepted for modelling TEs, as it attributes pupils' academic improvement to teacher unobserved characteristics or skills, conditional on the educational context (Hanushek (1986)).

Value-Added models (VAMs) have also been used for modelling "School Effects" (SEs), and the main concerns regarding validity and consistency of estimators are basically the same as when VAMs are used to estimate TEs (e.g. Reardon and Raudenbush (2009)).[1]

In the literature on SEs, the discussion has been mainly centred on the model specification and the reliability of SE estimates, which are used to construct League Tables (LTs) or school rankings. In particular, the fact that SE estimates are used to publish school rankings has made the discussion both more intense and more contentious. Apart from the uncertainty of SE measures that might cause misclassification problems in terms of schools ranking and difficulty to compare schools one-to-one, there are issues which are related to the negative consequences of LTs.

The related issues with LTs mostly converge on student selection problems which exacerbate student-teacher-school non-random assignment process. Pupils and teachers sorting between schools potentially affects less advantaged students and the schools which serve them. However, measures of school effectiveness are necessary for monitoring purposes, either privately or publicly as it also helps the implementation and the spread of better oriented polices. Hence, the discussion focuses on how well contextualised and reliable the SE estimates are, and how these estimates might be used.

Ideally, the research on school effectiveness should provide directions on how to improve the current quality of education, rather than inducing an oportunistic behaviour. In this context, we propose to analyse SE estimates and their stability over time in order to identify factors associated with highly effective schools.

To achieve this purpose and address uncertainty issues regarding SE estimates, we try to minimise the misclassification problem defining four main groups in school quality rankings:*High, Mid-High, Mid-Low, Low*. Although we still have problems of misclassification around the cutoffs, we are interested only in long run school performance, seeking to identify the highly effective (or less effective) schools in a period of time.

The analysis requires us to define a period of time over which it is possible to observe or track selected schools cohorts and consistently estimated SEs and TEs. Fortunately, our rich data from the Chilean school system, allow us to investigate school and teacher quality for a period of five years (2005 - 2009), from which we define a sample of schools that are trackable, given the feasibility of producing school and teacher Value-Added estimates. From this group of schools, which we call the reduced school panel (RSP), we can classify those schools which show

[1]In Chapter 2, we reviewed in detail VAM specifications and we discussed the main estimation approaches.

upward or downward trajectories in their effectiveness measures.

This approach for studying school effectiveness is novel in two aspects: (i) it reduces the risk of school quality misclassification when defining school rankings for a particular year and cohort; and (ii) the analysis of school quality transitions over time and the study of the key factors affecting them is founded on SEs and TEs, which is new to the literature, to our knowledge.

Our results suggest that neither downward nor upward trajectory of school effectiveness is explained by differences in school observed characteristics. The type of school dependence does not seem to determine the observed quality movements either. What drives the school effectiveness in the long run is the proportion of *High* (or *Low*) quality teachers based on our TE estimates.

6.2 Literature review

In the literature we observe that estimation of SEs has been used mainly to construct LTs, where schools are ranked based on their effectiveness at improving pupils' academic performance within a period of time. The model specifications have been generally presented as Hierarchical or Multilevel Models, which describe typical nested school organisation structure (e.g. Raudenbush and Bryk (1986); Goldstein (2011)). Alternatively, the so-called VAMs have put emphasis on the additional contribution, controlling for the educational context, that schools provide to students' achievement.

Hierarchical or Multilevel Linear Models (HLM) use the Random Effects (RE) approach where schools are assumed to be randomly drawn from the population distribution, representing the unobserved heterogeneity in the model (e.g. Skrondal and Rabe-Hesketh (2009)). On the other hand, VAMs refer just to the specification of the model, and they can be estimated either using a RE approach or a Fixed Effects (FE) approach which control for correlation between unobserved school heterogeneities and other covariates, including school dummy variables into the model.[2]

To compare schools adequately and arrive at valid inferences on school performance, it is necessary that effectiveness measures do not depend on the type of students attending a particular school. Raudenbush and Bryk (1986) consider a two level model where students, at first level, are nested in schools, at the second level. The SEs are assumed to be randomly distributed, and the differences in their point estimates represent the variability in effectiveness across schools. The point estimates of SE per school are obtained using the Empirical

[2]See Chapter 2 the discussion of VAM estimation approaches with explained examples of their applications.

Bayes (EB) approach which weights the regression coefficient from the within-school model at individual level and the estimated mean from the between-school model at school level. The weights represents how reliable the SE estimates are given the sampling error within schools. Thus, the SE estimates from schools with large sampling errors are shrunk towards the school population mean.

We use the EB methodology to estimate the posterior mean of the SE distribution, based on initial estimated parameters of covariates and sample variance distributions. Raudenbush and Bryk (1986) find that impacts of student socioeconomic background on pupils Maths achievements vary across schools. However, it is important to ensure that the results are not affected by student to school non-random assignment.

The issue of the student-to-school non-random assignment condition to estimate HLM is also addressed by Goldstein and Spiegelhalter (1996). The authors propose the default assumption of a homogenous school distribution, which states that if pupils attending schools are similar then schools can be considered as taken from the total population distribution. Therefore, predictions of SEs depend on the between-school variation and its adjusted EB distribution.

Reardon and Raudenbush (2009) also state that for causal school effects inference on pupil academic performance, random assignment of students-to-school is a required condition. Thus, to compare school effectiveness levels, it is assumed that every school sample is taken randomly from the total potential outcomes distribution.

In practice, it is impossible to construct the total potential outcome distribution because there are always limitations for students to attend particular sets of schools (e.g. due to distance, fee charges or other types of selection.). Therefore, the identifying assumption used to estimate SE parameters relies on the random assignment of students-to-schools.

Goldstein (2014) highlights the need to contextualise VAMs. SE estimates should not be driven by factors that do not depend on school and teacher ability to boost student academic performance. Therefore, it is required that school rankings are not determined by student-to-school non-random assignment due to the student selection process.

The author propose to estimate the VAM by the RE approach, where school unobserved heterogeneities and the error term are *iid* (independent and identically distributed). They should not be mutually correlated neither with respect to the other observable variables included in the model. The SE estimates obtained under this approach refer to the EB distribution of school random components.

In most of the papers discussed above, we observe the Multilevel or Value-Added Model specifications but not their derivation from a general achievement function. Hence, it is difficult to assess endogeneity problems if it is not clear what type of restrictions have been imposed to address model misspecification issues. In general, we do not find accurate discussions of potential endogeneity sources. For example, when VAMs include previous student scores, authors ignore the correlation between time-invariant unobserved individual effects and previous test scores. Under the RE approach, the estimators would be inconsistent.[3]

Differences in model specifications and the uncertainty of school effectiveness predictions due to relatively small samples of pupil observations per school, brings the discussion of how reliable school LTs are, and how precise the comparison between schools is. LTs have been used across different sectors apart from educational establishment, such as hospital institutions and other public services organisations.

Although there are positive aspects from school LTs, such as: improving control mechanisms and the allocation of resources due to better public accountability (Foley and Goldstein (2013); Goldstein (2014)), LTs also generate some negative spillovers.

In principle, LTs can reduce asymmetries of information providing well processed statistics data to users. However, this fact generates directly some unfavourable factors. If schools are publicly ranked, it is more likely that both good teachers and administrators are more reluctant to work for more disadvantaged schools, in terms of students and resources. Similar situations occur from the demand side, where parents might prefer high ranked schools, exacerbating problems of non-random assignment of students-to-schools and teachers-to-schools (Ladd and Walsh (2002); Foley and Goldstein (2013)).

Another important drawback from LTs is what is known in the literature as *gaming*, which refers to all discretionary actions taken by principals (or teachers) in order to improve their school's ranking. These actions may have harmful implications: schools focus just on improving test scores rather than providing a holistic education, sorting students into classes, and higher selection problems (Goldstein (2014)).

Leckie and Goldstein (2011) suggest the use of contextualised VAMs that must include a vast set of variables which allow one to account for initial differences between schools. If the VAM fails to control for original school differences, the LT could yield a misleading school effectiveness ranking. In most cases, SE estimations cannot be distinguished from the total mean if standard errors, which are of the point estimates of individual SE cross the mean. Authors present Value Added measures with 95% confidence intervals, showing how uncertain the school ranking

[3]The same problem occurs when the history of teacher effects (or teacher dummies) are included in the model, and their correlation with previous pupil scores is ignored.

is. This reliability factor depends on the number of observations per school used to estimate SEs, the smaller the sample the higher the uncertainty of estimates.

For their sample of secondary schools in England, Leckie and Goldstein (2011) argue that LTs are misleading in the middle of the ranking, as it is not possible to compare one-to-one schools significantly. However, SEs at the bottom and at the top of the distribution are significantly different from each other. Hence, a categorisation of schools into groups is an alternative to work with, although the concern would be then on the cutoff decision.

The uncertainty of school rankings based on SEs has opened the discussion about how to use or not to use LT results. Foley and Goldstein (2013) have suggested a couple of recommendations such as minimising the amount of reward provided to schools based on LTs in order to reduce the pressure put on principals, teachers and even students, for improving the performance on standardised exams. In addition, they propose LTs must be accompanied with a reliability measure of SE estimates. Furthermore, authors advise the use of SE estimates for programme implementation and evaluation rather than for comparing establishments individually.

Ideally, research on school effectiveness should provide directions on how to improve the current quality of education rather than inducing unproductive or harmful behaviour of principals, teachers and families.

6.3 School Effects: stability over time

Apart from the uncertainty issues concerning SE estimates and the difficulty of one-to-one comparisons in schools which are close in the distribution, Leckie and Goldstein (2011) expose an additional concern on English LTs. They state that SE estimates for one year are obtained from a specific cohort which was followed through all secondary grades within a particular context. However, by the time LT results are released and parents make decisions based on them, it does not necessarily mean that school effectiveness predictions would be relevant for pupils just starting the secondary school.

We believe that the long run stability of school quality levels is as important as the reliability of the yearly SE estimates, although we have not found studies researching that. A similar question is addressed by McCaffrey et al. (2009), but focusing on TEs stability over time. The authors find that teacher Value Added measures are relatively unstable, with a year-to-year correlation between 0.2-0.3. Moreover, the authors suggest that changes in the VAM specification do not significantly change the TE stability levels.[4]

[4] Note that we are not able to investigate the stability over time of teacher effect estimates

To investigate SEs stability, we propose working on school quality categories that divide SEs in quartiles. This broader classification of SEs solves the problem of uncertainty in the schools ranking comparisons. We allocate schools in school quality categories given the division of SE distribution in quartiles: *Low, Mid-Low, Mid-High* and *High*.

Although we still have the misclassification problem around the cutoff, we minimise its effects because we focus on the stability of the quality levels, and particularly its trajectory, rather than the single yearly classification. Hence, we are interested in schools' *upwards* or *downwards* trajectories, and we try to explain the factors shaping observed movements.

6.4 School effects (SEs) estimation

To estimate consistently SEs for consecutive cohorts, we face the same difficulties we had when we estimated TEs for the 4^{th} grade 2005 cohort. In Chapter 5, we have discussed the multilevel structure of the educational system and the implications for the identification strategy of school and teacher effects.

In theory, we require students randomly assigned to schools and teachers, and teachers randomly assigned to schools and classrooms. In practice, neither students nor teachers are likely to be randomly assigned to schools. In addition, a student might not be randomly assigned to a teacher or a set of teachers within a school, depending on the principal's or school's rules. Furthermore, teachers can be hired for one or more schools and be allocated to a single class or to several classes teaching a specific subject within a grade and school.

Similar to most available educational datasets, in our data we observe specialised tracks and teachers by area from 5^{th} grade onwards. This specialisation in subject complicates significantly our estimation strategy. Therefore, to reduce the non-random assignment issues we focus on 4^{th} grade cohorts taught by general teachers only.[5]

Additionally, as we are interested in estimating SEs and TEs simultaneously, we need schools with at least two 4^{th} grade classes per cohort, taught by two different general teachers. The details of the selection sample are discussed later on.

In Chapter 2, we derived four VAMs for which we set different parameter

presented in Chapter 5, because the group of teachers used to obtain Value Added measures do not teach every year in 4^{th} grade, when the Simce exam is taken. Besides, from those who stay teaching in 4^{th} grade, the sample selection required to identify teacher unobserved heterogeneities makes that some teacher observations disappear from the estimation sample.

[5]We understand as general teachers, those teachers who are fully assigned to a single classroom and teach at least the main two subjects Language and Maths.

restrictions on the rate of persistence λ, based on the values commonly observed in the literature. In previous Chapter 5, we estimated three of theses VAMs, and we concluded that **Modelo 3** which considers full persistence of input factor effects ($\lambda = 1$) is likely to be biased. Assuming $\lambda = 1$ may provide misleading results due to the omission of relevant variables such as, the lagged scores variable and the individual unobserved heterogeneity.

On the other hand, we found very similar estimated coefficients and teacher Value Added measures when estimating **Model 4** which preset a fixed value for $\lambda = \lambda_0$, and the unrestricted **Model 1** which freely estimates λ ignoring the potential endogeneity problem caused by the possible correlation of lagged scores and the error term.

We are aware that if **Model 1** is estimated following the RE approach the estimator would be inconsistent. Therefore, we propose **Model 4** to be our preferred VAM specification, pre-setting the rate of persistence λ equal to λ_0. Nevertheless, we estimate **Model 1** as well for comparison purposes.

In this current chapter, we focus exclusively on the teacher and school effect estimates obtained from the empirical Bayes distribution (EB) after estimating **Model 4** by the Maximum Likelihood estimator (MLE). This estimation methodology is supported by **Approach 3**, from Chapter 2.[6]

Model 4 ($\lambda = \lambda_0$):

$$A_{i,g} - \lambda_0 A_{i,g-1} = x'_{i,g}\beta_g + T'_{j,g}\pi_g + \tau_{j,g} + S'_g\theta_g + s_g + \alpha_i + \varepsilon_{i,g} - \varepsilon_{i,g-1} \tag{6.1}$$

where $A_{i,g}$ correspond to Language and Maths achievement. The vectors $x_{i,g}$, T_g and $S_{j,g}$ are sets of observable variables at student, teacher and school level. Additionally, α_i, $\tau_{j,g}$ and s_g correspond to unobserved part, and what we know as individual ability, teachers and school effects. The sub-index j identifies the teacher, and g the grade in which student i is enrolled in.

However, for estimation purposes we separate equation (6.1) in two achievement functions for Language and Maths, respectively. Both equations are simultaneously estimated imposing some restrictions on the persistence parameter λ_0 and the loading factors $\delta_{L,\alpha}, \delta_{M,\tau}, \delta_{L,g}$ in the equation below.

$$L_{i,4} - \lambda_0 L_{i,3} = x'_i\beta_L + T'_j\pi_L + S'\theta_M + 1.\alpha_i + \delta_{L,\tau}\tau_j + \delta_{L,s}s_g + \varepsilon_{i,L} \tag{6.2}$$

$$M_{i,4} - \lambda_0 M_{i,3} = x'_i\beta_M + T'_j\pi_M + S'\theta_M + \delta_{M,\alpha}\alpha_i + 1.\tau_j + 1.s_g + \varepsilon_{i,M} \tag{6.3}$$

where the subscripts 4 and 3 refer to the grades. M and L are the test scores

[6] We have discussed in previous Chapters 1 and 4, if there is random assignment of teacher to classroom and we are working with cross-section samples, the MLE-EB performs well in predicting teacher and school effects (Guarino et al. (2014a,b))

for Maths and Language exams, respectively. The correlation between the pairs of random effects in the two equations is captured using a factor loading representation as shown above. For identification purposes we impose the following normalisation restrictions, $\delta_{M,\tau} = \delta_{M,s} = \delta_{L,\alpha} = 1$.[7] The other three factor loadings ($\delta_{M,\alpha}, \delta_{L,\tau}$ and $\delta_{L,s}$) are freely estimated.

To consistently estimate equations (6.2) and (6.3) using the above MLE-EB method, we require to hold the following set of assumptions.

Assumption A1. Strict Exogeneity of covariates:

$$E[\varepsilon_{i,M}, \varepsilon_{i,L} | past,\ present \text{ and } future \text{ values of } (x, T, S), \alpha_i, \tau_j, s_g] = 0.$$

The above strict exogeneity assumption will fail, if for example the students are sorted into classes based on either observable or unobservable characteristics or both. This leads us to the second set of assumptions we require; ***Assumptions A2.1***, ***A2.2*** and ***A2.3***.

In case there is non-random assignment of students to teachers or classes, the strict exogeneity assumption will be violated. To hold assumption **A1** is necessary to impose three additional assumptions:

Assumption A2.1: Random assignment of students to schools.

Assumption A2.2: Random assignment of students to teachers (or teacher to classrooms).

Assumption A2.3: Random assignment of teachers to schools.

This set of assumptions, from ***Assumption A2.1*** to ***Assumption A2.3*** claim random assignment. Each assumption above is thought to deal with three different sources of endogeneity in the covariates.

For the first source of endogeneity we can only control for family background and student characteristics, as other authors have done (e.g. Rothstein (2009, 2010); Chetty et al. (2014a)). With respect to the second source of endogeneity, we have statistically tested in Chapter 4 whether there is evidence of non-random assignment of teacher to classrooms based on previous marks, and the results support our non-random assignment assumption. Lastly, for the third source of endogeneity, we are unable to prove non-random assignment of teacher to schools,

[7]The restriction where one of the loading factor of a latent variable is set equal to 1 is also known as the normalisation constraint or as the unit-loading rule (Rabe-Hesketh et al. (2004); Heckman et al. (2006)). The *loading factor* which is constrained to 1 is also called the *anchor* value for the corresponding latent variable. The inference made on the estimated *loading factor* is with respect to the normalised factor.

and we just assume this assumption holds.

However, we support assumption A2.1 and A2.2 based on the results obtained from Chapter 5 regarding the TE estimates under different types of school dependence samples. In this analysis, we found that the estimated teacher impact of *Municipal* teachers does not vary when the sample includes private schools, which are thought to be more likely to violate non-random assignment assumptions as they are allowed to select students and they have more flexibility to recruit teachers.

Apart from the sources of endogeneity due to observed covariates, we also need to impose a new set of assumptions in relation to the unobserved heterogeneities factors.

Assumption A3.1: $E[\alpha_i|past, present$ and $future$ values of $x] = 0$.

Assumption A3.2: $E[\tau_j|past, present$ and $future$ values of $T] = 0$.

Assumption A3.3: $E[s|past, present$ and $future$ values of $S] = 0$.

Note, under ***Assumption A2.1*** to***A2.3***, the requirements for ***Assumptions A3.1*** to ***Assumptions A3.3*** have become simpler. For example, if there is random sorting of students to teachers/classes and also of teachers to schools.

$$E[\alpha_i|past,\ present\,\text{and}\,future\,(x,T,S)] = E[\alpha_i|past,\ present\,\text{and}\,future\,x]$$

To estimate TEs and SEs in **Model 4**, following the RE approach, we estimate (6.2) and (6.3) simultaneously by MLE. Then, we obtain the EB distribution for teacher and school effects.

In order that MLE is a consistent estimator, we assume that unobserved heterogeneities are drawn from a random distribution.[8] If the above assumptions hold, MLE-EB will provide consistent estimators. Then, we require that all the random errors are *independent of each other*, and have the following distributions:

$\alpha_i \sim iid\ N(0,\sigma^2_\alpha)$; $\tau_j \sim iid\ N(0,\sigma^2_\tau)$; $s_g \sim iid\ N(0,\sigma^2_s)$; $\varepsilon_{i,M} \sim iid\ N(0,\sigma^2_\varepsilon)$; $\varepsilon_{i,L} \sim iid\ N(0,\sigma^2_\varepsilon)$.

Note, we restrict the variances to be the same for $\varepsilon_{i,M}$ and $\varepsilon_{i,L}$.[9]

Having established the estimation methodology and related assumptions, we now select the sample from our original cohorts that allow us to identify si-

[8] To estimate the model and obtain the EB estimates for the latent variables, we use the Generalised Structural Equation Modelling (GSEM) programme created by Rabe-Hesketh et al. (2004, 2007) and available in Stata 13 (StataCorp, 2013).

[9] Due to convergence problems, we could not freely estimate unrestricted error variances for $\varepsilon_{i,M}$ and $\varepsilon_{i,L}$.

multaneously school and teacher effects. Hence, in this chapter we propose both: a broader classification for teacher and school effects and and a long run analysis of their measures.

6.4.1 Sample selection

We have already discussed how the estimation of TEs and SEs require specific conditions on the sample, particularly if we are now interested in the stability of SEs. Similar to the selection made for Chapter 5, we take a sample from the cross section cohorts from 2005 to 2009.

Initially, we select the same 4^{th} grade cohort used in Chapter 5 to set up the baseline for teacher and school effects. Therefore, we impose the same requirements on the sample, with schools having at least two 4^{th} grade classrooms taught by full-time general teachers. In addition, the most relevant variables for student and teacher characteristics must be available. The rest of the selection conditions are explained later.

Firstly, we focus only on early primary grades because we want to avoid estimation difficulties due to the tracking of courses by field, that also brings in issues concerning teacher specialisation by area. That is not the only technical estimation issue, but it means new sources of endogeneities.

We do not continue the analysis beyond 2009 onwards because the number of schools switching to subject specialist (SS) teacher scheme, even in 4^{th} grade cohorts, increases significantly. This trajectory has been observed since 2006, but in 2010, our student sample would be reduced dramatically under this criterion, affecting the number of school that we could track in our analysis. In the next tables, we show the weight of each selection criterion for the whole selection sample.

The characteristics of the Simce exam not only restricts us in terms of grades and cohorts that are available for estimations, but they also determine the model specification and estimation strategy. In our particular case, the grades and cohorts chosen to estimate TEs and SEs only have National examination for 4^{th} grades. Our VAM, presented above in equations (6.2) and(6.3), require an achievement measure for 3^{rd} grade. We use Language and Maths school marks (standardised at school level) as a proxy for unobserved 3^{rd} grade Simce scores.[10] We use the equivalent proxy for every 4^{th} grade cohort from 2005 to 2009.

We maintain the same selection criteria for all cohorts, and we show in Table 6.1 the variablility from year to year. The selection process follows a particular order which is classified in three main groups: (i) selection related to teacher

[10]In Chapter 3, we checked the correlation between standardised Simce scores and standardised school marks to support the use 3^{rd} grade school marks as a proxy for unobserved standardised National examination scores.

conditions, (ii) selection with respect to the availability of information at student and level, and (iii) selection at school level conditions.

Table 6.1: Sample selection criteria
4^{th} grade cohorts (2005 - 2009)

	Total observations original bases:		**2005**		**2006**		**2007**		**2008**		**2009**	
			268,162	**100%**	**265,681**	**100%**	**257,344**	**100%**	**254,673**	**100%**	**250,275**	**100%**
Sample selection			**Dropped 2005**	**%**	**Dropped 2006**	**%**	**Dropped 2007**	**%**	**Dropped 2008**	**%**	**Dropped 2009**	**%**
Teacher selection	(1) Dropping obs. without teacher ID assigned	Yes	3,706	1%	3,787	1%	3,080	1%	4,667	2%	1,281	1%
	(2) Dropping obs. with teachers observed in 2 o more schools	Yes	23,973	9%	21,070	8%	17,672	7%	13,441	5%	12,567	5%
	(3) Dropping obs. with specialised subject teachers	**Yes**	**37,458**	**14%**	**44,294**	**17%**	**49,782**	**19%**	**59,176**	**23%**	**71,111**	**28%**
	(4) Dropping obs. without teachers characteristics	Yes	14,671	5%	12,720	5%	15,467	6%	14,061	6%	11,734	5%
Student selection	(5) Dropping obs. without Simce Exam	Yes	13,005	5%	15,415	6%	12,408	5%	12,174	5%	17,752	7%
	(6) Dropping obs. with repeating students (Not observed in 3rd grad	Yes	8,390	3%	8,030	3%	7,315	3%	6,506	3%	5,622	2%
	(7) Dropping obs. without School Marks - 4th grade	Yes	302	0%	262	0%	177	0%	158	0%	41	0%
	(8) Dropping obs. without School Marks - 3rd grade	Yes	869	0%	437	0%	341	0%	485	0%	799	0%
	(9) Dropping students classified as Special Needs	Yes	840	0%	962	0%	722	0%	986	0%	296	0%
	(10) Dropping obs. without student characteristics	**Yes**	**19,555**	**7%**	**11,966**	**5%**	**35,326**	**13%**	**12,387**	**5%**	**20,930**	**8%**
School selection	(11) Dropping obs. without school characteristics	Yes	-	0%	-	0%	-	0%	-	0%	-	0%
	(12) Dropping obs. with class letter "F" label	Yes	305	0%	376	0%	318	0%	306	0%	240	0%
	(13) Dropping classes without a minimum No. of students per class	15	13,462	5%	12,366	5%	11,140	4%	13,014	5%	14,702	6%
	(14) Dropping classes without a maximum No. of students per class	45	-	0%	647	0%	46	0%	92	0%	46	0%
	(15) Dropping schools without a minimum of classes per grade	**2 or more**	**41,429**	**15%**	**42,482**	**16%**	**39,725**	**15%**	**41,996**	**16%**	**37,205**	**15%**
	(16) Dropping schools with less general teachers than classes per gra	Yes	919	0%	1,606	1%	247	0%	1,399	1%	1,372	1%
	Total observations selected samples:		**2005**		**2006**		**2007**		**2008**		**2009**	
			89,278	**33%**	**89,261**	**34%**	**63,578**	**25%**	**73,825**	**29%**	**54,577**	**22%**

Notes: **(i)** We dropped observations sequentially from the original bases: "4^{th} grade 2005-09" cohorts. **(ii)** The first part of the selection sample we show the number of observations eliminated given individual conditions. **(iii)** The second part, after the dividing line, we dropped observations based on classroom and school specification. **(iv)** We eliminate pupils attending classrooms labeled as F in row (4) because this was our classification to all classes which the letter assigned in the original database was not between A and E, and that might represent a mistake. **(v)** The student characteristics in row (10) refer to mother's education, and household income.

The first selection criteria group mainly drops observations because of the presence of teachers specialised in subject or because teachers were teaching in more than one school during the same year. We see how the proportion of pupils being taught by SS teachers increases from 14% in 2005 to 28% in 2009, while the proportion of teachers observed in two or more schools actually decrease from 9% to 5% during the same period (see rows (2) and (3), Table 6.1).

The second selection criteria group refers to students. Here we require that every student in the sample has taken the Simce in 4^{th} grade and also has available their school marks in 3^{rd} grade for the previous year. We also eliminate students from the sample who have been in 4^{th} grade but whom we do not observe in 3^{rd} grade the year before. Nevertheless, the most important source of selection is with respect to individual background, which is obtained from the parental questionnaires when the Simce is taken. The proportion of students generally dropped here is between 5% and 8%, but in 2007 this proportion increases up to 13%. However, this is a particular problem with the original dataset in 2007, where we were not able to match some of the student IDs with their parental Simce questionnaire. Thus, we had to increase the number of observations dropped from the sample in this year.

Regarding the selection criteria at school level, we follow the literature and keep only classrooms with at least 15 students per grade. Here, we want to reduce

TE estimation uncertainty for those teachers who teach in very small classes. On the other hand, we also eliminate classrooms with more than 45 students per class because of likely misclassification, as the maximum allowed by law is 45. The most important source of selection at the school level is the condition that there are at least two classes per 4^{th} grade, which generally represents around the 15% of the total sample.[11]

In comparison with the sample selection employed for the estimation of TEs in Chapter 5, here we add the restriction of dropping students classified as "Special Needs". However, the estimation for TEs and SEs do not vary significantly with this extra condition and we prefer to hold it in the current chapter.

Table 6.2 shows an alternative way to present the selection in terms of teachers and schools dropped after the selection criteria is applied at teacher, student and school level. It is important to highlight that we follow an order of selection, so the number of observations dropped in every row is in relation to the remaining data after the previous selection criterion. This means that some of the observations already dropped in earlier conditions (e.g. Teacher selection) may have also been eliminated by later conditions.

In the Table 6.2 we observe that every year the amount and proportion of observations dropped from the original sample because of teacher selection conditions increased considerably from 2005 to 2009. In 2005 the percentage of teachers and schools eliminated for this criterion was 25% and 16%, respectively, increasing up to 47% and 38% in 2009. In contrast, the percentage of observations dropped because of student and school selection criteria stays relatively stable over the period.

[11]See Appendix 2.1, the table with a summary of the dataset mostly used in this literature and some of their sample selection process.

Table 6.2: Number of observations dropped by type of selection
4^{th} grade cohorts (2005-2009)

		2005		2006		2007		2008		2009	
Original 4th grade cohort						**Before selection**					
	Number of Students	268,162	100%	265,681	100%	257,344	100%	254,673	100%	250,275	100%
	Number of Teachers	12,233	100%	12,166	100%	12,052	100%	12,052	100%	12,245	100%
	Number of Schools	8,338	100%	8,267	100%	8,180	100%	8,180	100%	8,164	100%
						Dropped observations					
Teacher	Number of Students	79,808	30%	81,871	31%	86,001	33%	91,345	36%	96,693	39%
selection	Number of Teachers	3,094	25%	6,809	56%	5,119	42%	5,286	44%	5,715	47%
	Number of Schools	1,355	16%	1,004	12%	2,990	37%	3,011	37%	3,125	38%
						Dropped observations					
Student	Number of Students	33,150	12%	37,072	14%	56,289	22%	32,966	13%	45,438	18%
selection	Number of Teachers	1,092	9%	481	4%	1,021	8%	24	0%	489	4%
	Number of Schools	1,074	13%	530	6%	747	9%	280	3%	371	5%
						Dropped observations					
School	Number of Students	65,926	25%	57,477	22%	51,476	20%	56,537	22%	53,567	21%
selection	Number of Teachers	4,907	40%	1,831	15%	3,726	31%	4,155	34%	4,038	33%
	Number of Schools	4,576	55%	5,421	66%	3,492	43%	3,766	46%	3,790	46%
Selected 4th grade chort						**After selection**					
	Number of Students	89,278	33%	89,261	34%	63,578	25%	73,825	29%	54,577	22%
	Number of Teachers	3,140	26%	3,045	25%	2,186	18%	2,587	21%	2,003	16%
	Number of Schools	1,333	16%	1,312	16%	951	12%	1,123	14%	878	11%

Note: The first section shows the number of students, teachers and school observed in the original cohorts. The three selection groups show the amount of student, teacher and schools dropped from the original cohort because of the selection criteria with respect to their conditions. In the last group, "Selected 4^{th} grade cohort" we show the final number of students, teacher and schools that we use in our sample to estimate both teacher and school effects.

The sample selection criteria applied to all 4^{th} grade cohorts from 2005 to 2009 generate samples that do not differ considerably from their original cohorts. The main differences between the samples and the original cohorts are given by pupil achievement. From Tables 8.16 to 8.20, in Appendix 6.1, we observe that both school marks and Simce scores are always slightly higher in the selected samples than the original cohorts. Other differences can be found in terms of classes per grade and the number of students per grade, which are direct consequence of our selection criteria. Regarding the changes in standardised Simce scores before and after selection, we also show in Appendix 6.1 from Figure 8.20 to Figure 8.24, that for the whole period the standardised Simce distributions of selected samples are slightly shifted to the right with respect to the original cohorts. However, the shapes of the distributions are very similar in both subjects, every year.

6.5 Results

The results presented in this section correspond to the yearly estimation of TEs and SEs using consecutive 4^{th} grade student student cohorts from 2005 to 2009. Technically, to estimate our VAM specified as **Model 4**, we require to preset a value for λ_0, which we have chosen to be equal 0.4, following the literature. In a first stage, we estimate equations (6.2) and (6.3) simultaneously by MLE, where we recover the estimated coefficients, including the *loading factors* and prior distributions of latent variables (students, teacher and schools unobserved heterogeneities). The second step consists of obtaining the EB distribution for each latent variable.

Additionally, we also estimate **Model 1** using the MLE-EB methodology, but only to be used as a reference. We are aware of the possible inconsistency of the MLE estimator in this model, given the likely correlation between lagged pupil scores and the error term.

We are particularly interested in the TE and SE estimations, which are the expected value of the EB distributions per teacher and school, respectively. Every year, we classify both teachers and schools in four quality groups: *High, Mid-High, Mid-Low*, and *Low*. The groups are formed by quartile distributions of TEs and SEs, where the High quality group is formed with those teachers and and schools with the highest predicted teacher and school effect, respectively.

In Table 6.3 we presented the estimated coefficients by MLE of **Model 4** ($\lambda_0 = 0.4$) and **Model 1** (*estimated* λ), for all 4^{th} grade cohorts from 2005 to 2009.

The first part of Table 6.3 shows the *factor loadings* estimates. Taking the unit-loading factor or *anchor factor* as pivot, we interpret the other estimated *factor loading* as the effect of the latent variable in the equation, relative to the pivot. For example, when the *anchor factor* is set for the impact of SEs on Language scores, it means that if its corresponding estimated *factor loading* is lower than 1, the SE on Maths is smaller than on Language. When the estimated *factor loading* is greater than 1, the impact in the dependent variable is larger with respect to the *anchor factor*.

The estimated coefficients of *factors loading* are very stable along the period, confirming the higher impact of SEs on Language scores with respect to Maths scores, and the higher impact of TEs on Maths scores with respect to Language. All estimated *loading factors* are statistically significantly different from 1 (or from the *anchor factor*). It seems there are no noticeable differences of estimated *loading factors* between **Model 4** and **Model 1**.

Despite we are aware of the consistency problems, we estimate **Model 1** by MLE, and we obtain estimated persistence parameters $\lambda_L, \lambda_M,$. They are very stable across years, around 0.51 and 0.54 for Language and Maths, respectively, and they are not very far from our preset value of $\lambda_0, = 0.4$ imposed in **Model 4**.

The interpretation of school and teacher effect predictions can be shown more intuitively in terms of expected changes in pupils achievement. If a student from the median distribution of Simce scores is exposed to 1 SD higher effective teacher or school, the impact on the standardised Simce scores is given by the estimated SD of the TEs and SEs, respectively, obtained from the EB distributions.

Table 6.3: Maximum Likelihood coefficient estimates
4^{th} grade selected sample cohorts 2005-2009

	2005				2006				2007				2008				2009			
	Lambda preset (λ=0.4)		*Lambda estimated*		*Lambda preset (λ=0.4)*		*Lambda estimated*		*Lambda preset (λ=0.4)*		*Lambda estimated*		*Lambda preset (λ=0.4)*		*Lambda estimated*		*Lambda preset (λ=0.4)*		*Lambda estimated*	
	Model 4		**Model 1**		**Model 4**		**Model 1**		**Model 4**		**Model 1**		**Model 4**		**Model 1**		**Model 4**		**Model 1**	
	Lang. (1)	Maths (2)	Lang. (3)	Maths (4)	Lang. (5)	Maths (6)	Lang. (7)	Maths (8)	Lang. (9)	Maths (10)	Lang. (11)	Maths (12)	Lang. (13)	Maths	Lang. (15)	Maths	Lang. (17)	Maths	Lang. (19)	Maths
Loading factors																				
School Effects	**1**	0.868***	**1**	0.869***	**1**	0.883***	**1**	0.868***	**1**	0.899***	**1**	0.894***	**1**	0.912***	**1**	0.910***	**1**	0.961***	**1**	0.948***
		(0.015)		(0.014)		(0.017)		(0.017)		(0.019)		(0.019)		(0.018)		(0.018)		(0.020)		(0.020)
Teacher Effects	**1**	1.467***	**1**	1.521***	**1**	1.496***	**1**	1.558***	**1**	1.572***	**1**	1.626***	**1**	1.687***	**1**	1.772***	**1**	1.870***	**1**	1.951***
		(0.022)		(0.024)		(0.022)		(0.024)		(0.031)		(0.033)		(0.034)		(0.037)		(0.055)		(0.058)
Student Ability	1.111***	**1**	1.134***	**1**	1.210***	**1**	1.245***	**1**	1.157***	**1**	1.178***	**1**	1.252***	**1**	1.295***	**1**	1.268***	**1**	1.318***	**1**
	(0.006)		(0.006)		(0.007)		(0.007)		(0.007)		(0.007)		(0.007)		(0.008)		(0.009)		(0.010)	
Student covariates																				
Previous Stdzd. Language School Marks			0.519***				0.513***				0.513***				0.514***				0.508***	
			(0.002)				(0.003)				(0.003)				(0.003)				(0.003)	
Previous Stdzd. Maths School Marks				0.545***				0.552***				0.539***				0.551***				0.557***
				(0.002)				(0.002)				(0.003)				(0.002)				(0.003)
Gender (Female=1)	0.020***	-0.067***	0.006	-0.060***	0.033***	-0.088***	0.020***	-0.080***	0.030***	-0.077***	0.018***	-0.071***	0.113***	-0.089***	0.099***	-0.084***	0.080***	-0.066***	0.065***	-0.058***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Mother education level	0.032***	0.029***	0.027***	0.023***	0.036***	0.043***	0.030***	0.035***	0.033***	0.035***	0.029***	0.030***	0.027***	0.037***	0.021***	0.030***	0.022***	0.031***	0.017***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Household income	0.019***	0.021***	0.016***	0.017***	0.012***	0.011***	0.010***	0.008***	0.013***	0.016***	0.011***	0.013***	0.013***	0.016***	0.011***	0.013***	0.016***	0.017***	0.014***	0.013***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
Class and Teacher covariates																				
Class size	0.015***	0.016***	0.013***	0.014***	0.015***	0.018***	0.013***	0.016***	0.015***	0.019***	0.013***	0.018***	0.012***	0.016***	0.011***	0.014***	0.015***	0.018***	0.014***	0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Peers average GPA	-0.512***	-0.609***	-0.332***	-0.383***	-0.601***	-0.785***	-0.336***	-0.445***	-0.408***	-0.586***	-0.252***	-0.382***	-0.364***	-0.574***	-0.197***	-0.341***	-0.300***	-0.465***	-0.156***	-0.256***
	(0.035)	(0.041)	(0.026)	(0.031)	(0.043)	(0.052)	(0.029)	(0.035)	(0.038)	(0.046)	(0.030)	(0.037)	(0.033)	(0.042)	(0.026)	(0.033)	(0.035)	(0.044)	(0.028)	(0.036)
Gender (Female=1)	0.087***	0.076***	0.075***	0.062***	0.119***	0.124***	0.105***	0.105***	0.033	0.070**	0.034*	0.073***	0.106***	0.119***	0.102***	0.113***	0.041**	0.065**	0.034*	0.056**
	(0.019)	(0.023)	(0.017)	(0.022)	(0.021)	(0.026)	(0.018)	(0.024)	(0.022)	(0.029)	(0.021)	(0.027)	(0.019)	(0.026)	(0.018)	(0.024)	(0.019)	(0.026)	(0.018)	(0.025)
Years of experience in the education system	0.004***	0.004***	0.004***	0.004***	0.003***	0.003***	0.003***	0.003***	0.002**	0.002**	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Teaching ratio (Hrs teaching / Hrs contract)	-0.068	-0.104*	-0.058	-0.097*	-0.079	-0.059	-0.085*	-0.065	0.051	0.058	0.049	0.051	0.048	0.075	0.042	0.063	-0.002	-0.007	-0.003	-0.015
	(0.049)	(0.055)	(0.046)	(0.051)	(0.052)	(0.061)	(0.047)	(0.055)	(0.059)	(0.070)	(0.056)	(0.066)	(0.051)	(0.062)	(0.048)	(0.058)	(0.050)	(0.062)	(0.048)	(0.059)
School covariates																				
Private voucher school	0.423***	0.421***	0.415***	0.409***	0.382***	0.431***	0.361***	0.405***	0.380***	0.415***	0.372***	0.404***	0.398***	0.456***	0.394***	0.447***	0.333***	0.380***	0.335***	0.379***
	(0.023)	(0.024)	(0.022)	(0.022)	(0.024)	(0.026)	(0.022)	(0.023)	(0.027)	(0.029)	(0.026)	(0.027)	(0.024)	(0.026)	(0.023)	(0.025)	(0.026)	(0.029)	(0.025)	(0.028)
Private unsubsidised school	1.124***	1.215***	1.091***	1.178***	1.167***	1.340***	1.083***	1.242***	1.111***	1.296***	1.073***	1.247***	1.105***	1.367***	1.066***	1.317***	0.995***	1.258***	0.974***	1.233***
	(0.052)	(0.053)	(0.049)	(0.049)	(0.050)	(0.054)	(0.045)	(0.046)	(0.057)	(0.060)	(0.054)	(0.057)	(0.049)	(0.053)	(0.047)	(0.049)	(0.052)	(0.057)	(0.050)	(0.053)
Rurality (Rural=1)	0.030	0.002	0.016	-0.015	-0.053	-0.087	-0.046	-0.077	-0.096	-0.077	-0.098	-0.080	-0.022	-0.012	-0.035	-0.031	-0.003	-0.030	-0.006	-0.034
	(0.054)	(0.055)	(0.052)	(0.052)	(0.056)	(0.059)	(0.052)	(0.054)	(0.074)	(0.078)	(0.071)	(0.074)	(0.056)	(0.061)	(0.054)	(0.058)	(0.068)	(0.075)	(0.066)	(0.072)
Constant	2.129***	2.771***	1.131***	1.503***	2.602***	3.594***	1.146***	1.705***	1.462***	2.348***	0.612***	1.219***	1.175***	2.315***	0.272*	1.042***	0.894***	1.799***	0.101	0.627***
	(0.209)	(0.245)	(0.160)	(0.189)	(0.251)	(0.304)	(0.172)	(0.208)	(0.226)	(0.274)	(0.185)	(0.225)	(0.193)	(0.243)	(0.156)	(0.197)	(0.205)	(0.259)	(0.171)	(0.216)
Prior SD of School Effects: σ_s	0.316***	0.274***	0.308***	0.267***	0.315***	0.278***	0.297***	0.257***	0.316***	0.284***	0.308***	0.276***	0.308***	0.281***	0.303***	0.276***	0.316***	0.304***	0.307***	0.291***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Prior SD of Teacher Effects: σ_τ	0.237***	0.347***	0.210***	0.320***	0.259***	0.387***	0.224***	0.348***	0.226***	0.355***	0.202***	0.329***	0.210***	0.354***	0.184***	0.327***	0.170***	0.318***	0.155***	0.302***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.005)	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	-0.003	-0.003	(0.002)	(0.002)
Prior SD of Individual Ability: σ_α	0.529***	0.476***	0.514***	0.454***	0.540***	0.446***	0.484***	0.423***	0.544***	0.470***	0.529***	0.449***	0.547***	0.437***	0.534***	0.412***	0.556***	0.438***	0.542***	0.411***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Error variance: σ_ε^2	2.129***	2.771***	0.201***	0.201***	0.212***	0.212***	0.216***	0.216***	0.191***	0.191***	0.197***	0.197***	0.195***	0.195***	0.200***	0.200***	0.204***	0.204***	0.207***	0.207***
	(0.209)	(0.245)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log likelihood		-168899.5		-167876.6		-172480.4		-171334.6		-119750.7		-119099.8		-138637.2		-137655.1		-103777.4		-102837.8
Number of Observations		89,256		89,256		89,232		89,232		63,563		63,563		73,806		73,806		54,510		54,510

Notes: (i) To estimate the models and obtain the Posterior SD of unobserved heterogeneities, we use the Generalised Structural Equation Modelling (GSEM) programme, available in Stata 13 (StataCorp, 2013). **(ii)** Standard errors in parentheses. **(iii)** *** $p<0.001$; ** $p<0.05$; * $p<0.01$. **(iv)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(v)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600. **(vi)** Both Mother education level and Household income variables are assumed to be the same the year reported in the Simce Parents questionnaire and the year before, for which we do not have records. **(vii)** The base category for the included dummy variables is shown next in brackets: Pupil's gender (Male); Special needs pupils (all the rest pupils); Teacher's gender (Male); *Private voucher* schools (Municipal and Unsubsidised Private school); *Unsubsidised Private* school (Municipal and Private voucher schools); Rurality (Urban area).

However, we are not particularly interested in the precise impacts of teacher and school effects. In this chapter, we are interested in the stability of SE and TE estimates in the policy context of their use as valid measures of school and teacher quality. For more details regarding the predicted impact on the median pupil percentile ranking when exposed to a highly effective teacher or a more effective school, see Tables 8.22 and 8.23, respectively, in Appendix 6.2.[12]

Although we have estimates of SEs and TEs for 5 years, they do not correspond to the same teachers or even the same schools. The frequency of Simce exams which is only taken in 4^{th} grade for early primary grades, and the type of teacher specialisation, make of more difficult to track individual TE estimates. In our data, we observe some general teachers being allocated either to a specific cohort, or being constantly assigned to the same grade. For the first group of specialisation we are unable to estimate TEs every year, while for the second we have more chances to observe those teachers within the selected school samples. With respect to school quality measures, we could potentially observe SE estimates over the period, but due to our constrained selection sample, there is a significant number of schools which are not observed consecutively for the whole period (see Tables 6.1 and 6.2 above).

For every cohort, we construct rankings of teachers and schools based on their TE and SE, respectively. As we want to study the stability of SEs, we have to select a group of schools which are trackable across the panel. Then, every establishment from what we call the reduced school panel (RSP) has an associated SE estimate and an average TE in every year. With these measures and the yearly rankings constructed from the original sample, we have a school quality classification and an average teacher quality category per school.

The next section shows a descriptive analysis of the type of teachers observed in the selected sample cohorts, in terms of teacher specialisation and teacher quality. In addition, we describe the selection of trackable schools, or the RSP, which is used to define three types of school quality movements, based on the school quality classification along the period. In the subsequent section, we study the factors which might determine the trajectories in school quality over time, in relation to the types of teacher composition within schools.

[12]In Appendix 6.2, we show the Simce scores distribution for the whole period in absolute and standardised terms. Besides, we present the EB distributions of each latent variable and two summary tables with the expected changes in the percentile ranking of an average pupil when she is exposed to 1 SD higher effective teacher or school, assuming everything else constant. Additionally, in Appendix 6.3, from Tables 8.24 to 8.28, we show the EB distributions for all latent variables (student, teacher and school heterogeneities).

6.6 Types of teacher and school description

The main purpose of the estimation results presented in the previous section is to obtain SEs and TEs as measures of school and teacher quality, respectively. Both estimates can be interpreted as quality indicators of primary schools and teachers at national level. Therefore, we have measures of school and teacher quality for five consecutive 4^{th} grade cohorts, between 2005 and 2009. The objective of this chapter is to study school quality stability and analyse the nature of correlation between transitions and the type of teachers and teacher quality.

Regarding the school quality, we construct a school ranking every year based on the estimated SEs, where schools in the top quartile (highest SEs) are defined as *High* quality schools, while those in the bottom quartile are identified as *Low* quality schools. Schools in the middle of the distribution are separated in *Mid-High* and *Mid-Low* school quality categories. Analogously, a yearly teachers' ranking is based on TE estimates, and they are also classified as: *High*, *Mid-High*, *Mid-Low* and *Low* quality teachers.

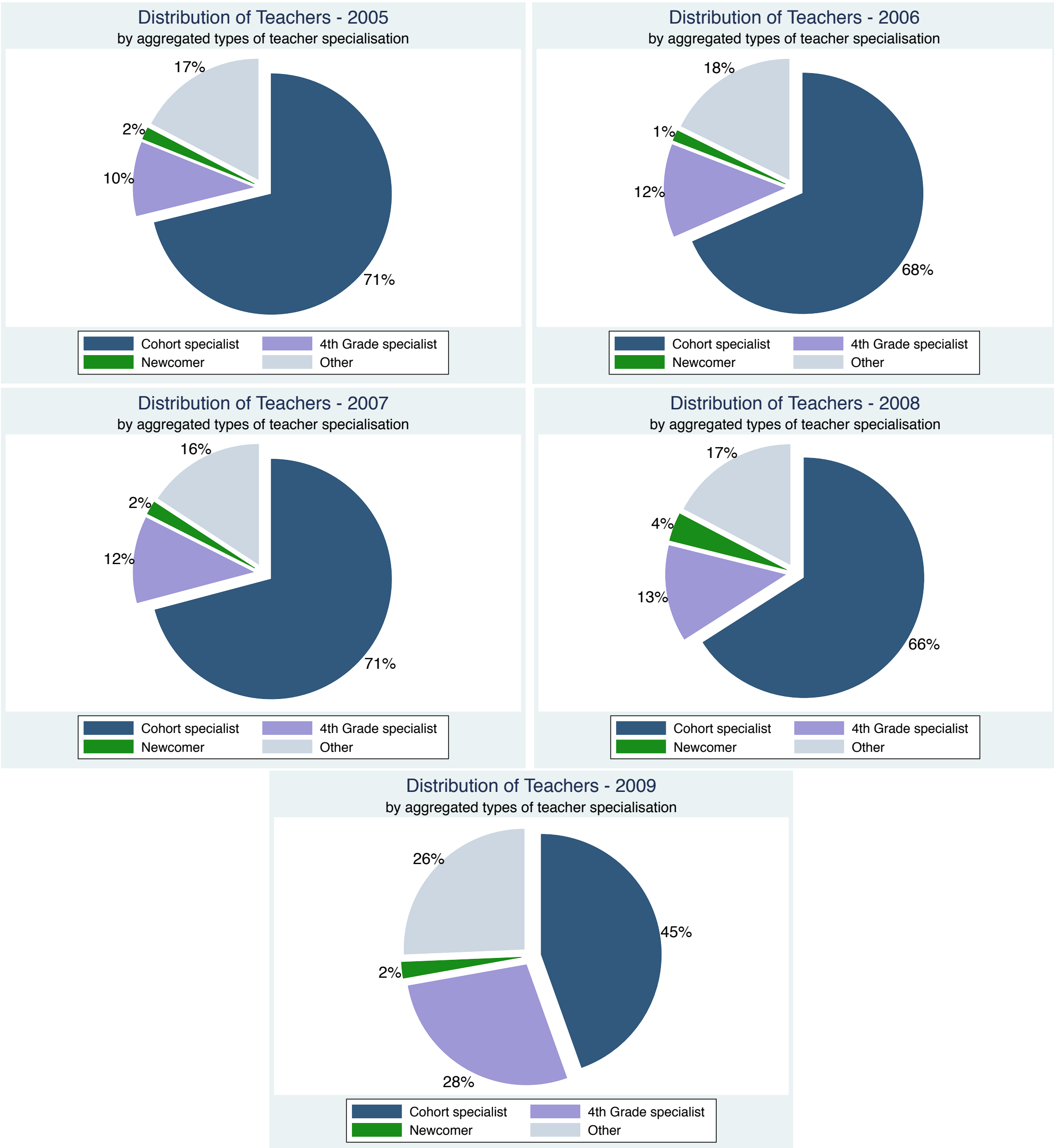
We are able to specify a school ranking only for those schools and teachers that satisfy the sample selection conditions required to estimate SEs and TEs. These restrictions produce a significant school attrition in our selected samples during the period 2005-2009. In later subsections we present a detailed description of school patterns, and the selection of the RSP which allow us to track schools and analyse their school quality evolution.

Individual TEs are more difficult to track as the Simce is taken only in 4^{th} and 8^{th} grade for primary school, and most of the early primary teachers who have taught in a 4^{th} grade return to lower grades (1^{st}, 2^{nd} and 3^{rd}). Some of the teachers follow the same class-cohort from early grades until 4^{th} grade when the Simce exam is taken. In the next subsection we show a descriptive distribution of teachers with respect to their type of training and their quality distribution.

6.6.1 Type of teacher specialisation

Taking advantage of the administrative dataset, we are able to track 4^{th} grade teachers for their previous year's activity. Hence, we identified in which school, grade and classroom they taught. We classify teachers in groups depending on their previous experience. The general classification is composed of: (i) **Cohort specialist** (CS), where general teachers were observed teaching the same class-cohort from the 1^{st}, 2^{nd} or 3^{rd} grade; (ii) **Grade specialist** (GS), where teachers have been teaching 4^{th} grade classes for the last two years at least, no matter the school; (iii) **Newcomers**, are those teachers who were not observed in previous years and have at most one year of experience in the system, (iv) **Others**, correspond to teachers who are not classified in any of the categories above (i.e. teachers who were teaching other grades and classrooms in the same or different school).

Figure 6.1: Type of teacher specialisation 2005 - 2009



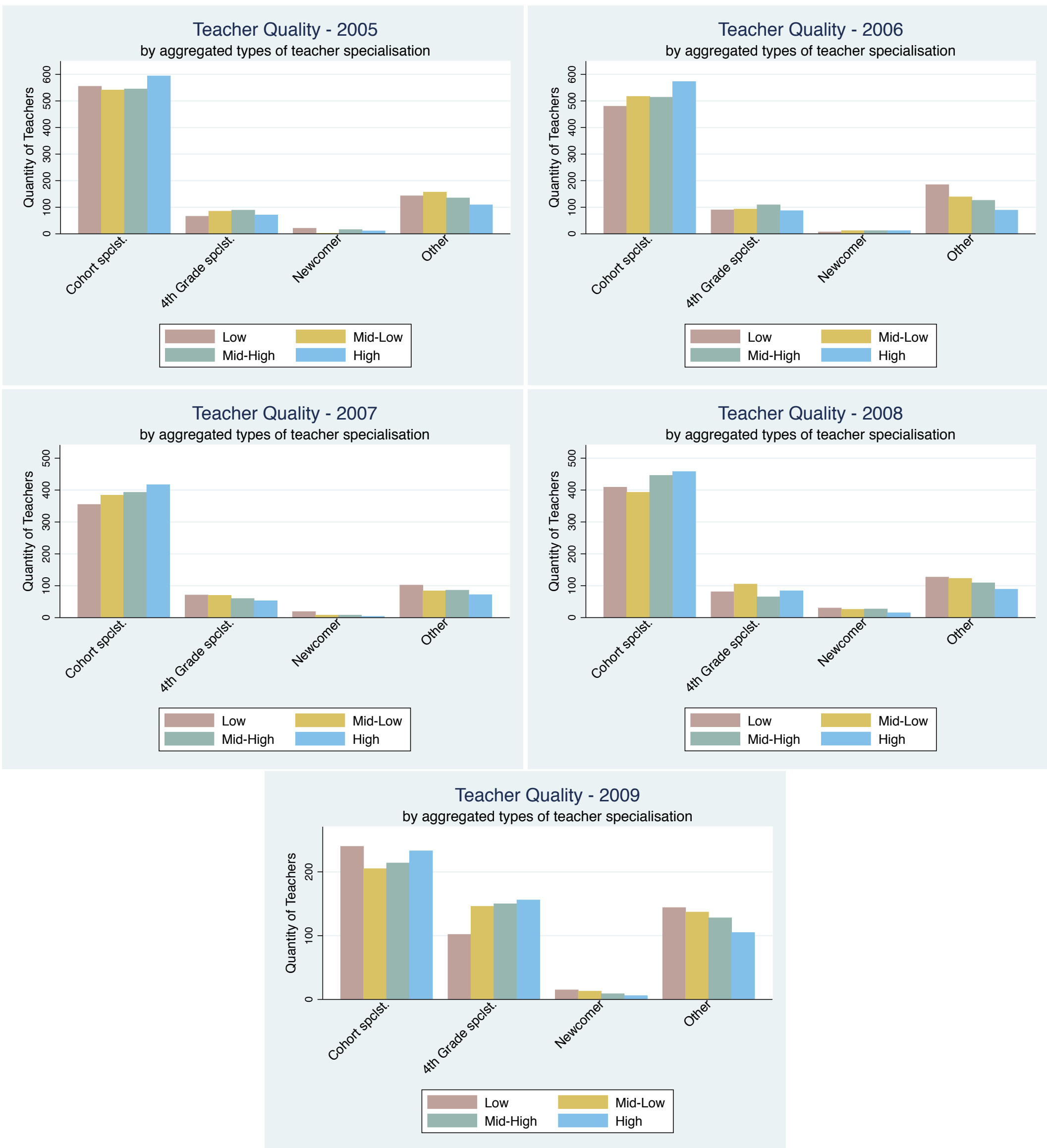
In Figure 6.1 we observe how from 2005 to 2008 the percentage of **Cohort specialist** (CS) teachers was around 70%, while the percentage of **Grade specialist** (GS) teachers was just 12% in average. The distribution dramatically changed in 2009, where the CS decrease to only 45% of the total teachers and the GS boost up to 28% of them. This change is observed across type of school dependence, although it is more evident in private schools, specially in *Unsubsidised Private* schools.[13]

[13]See Figure 8.25, Appendix 6.3, the type of teacher training distribution disaggregated by

It is important to highlight that there is a positive growth of subject specialist (SS) teachers over this period, and particularly in 2009 when the proportion almost doubled the number of students taught by SS teachers compared to 2005.[14] That implies we have to remove a higher proportion of teachers from the original cohort. Therefore, we observe that there is not only a positive movement to higher SS but there is also more specialisation on grades where general teachers are teaching in our sample.

Regarding the quality level of teachers and its distribution among the type of teacher specialisation, we observe in Figure 6.2 how there is no clear concentration of high quality teachers in any particular group, although there seems to be a majority within the group of CS teachers.

Figure 6.2: Type of teacher quality 2005 - 2009



groups within classroom and grade specialisation.

[14]See row 3, Table 6.1 in the Selection sample subsection.

In Appendix 6.3, we also show the disaggregated distribution of type of teacher training by quality level, suggesting a higher proportion of high quality teachers in cohort specialised categories from either 1^{st} or 2^{nd} grade, at least until 2008 (see Figure 8.29).

All this information is collapsed to school level, where we calculate the proportion of every type of teacher by quality category and type of training. In the next subsection, we present a detailed description of how we select the RSP from which we are able to analyse school quality stability and relate it to the type of teacher composition within school.

6.6.2 Trackable schools selection

The conditions imposed on the original cohorts in order to find estimates of SEs and TEs lead us to very restricted samples during this period. Between 2005 and 2009 we observe 1,944 schools in total, but only 15% of these appeared in all five years, as shown in the table below.

Table 6.4: Observed patterns in original school panel

Pattern	Fequency	Percent
11111	296	15%
1111X	167	9%
1XXXX	157	8%
11XXX	128	7%
11X11	126	6%
111XX	97	5%
X1XXX	89	5%
11X1X	89	5%
XXXX1	65	3%
Other	734	38%
Total	**1,948**	**100%**

Notes: **(i)** The original school panel has 1,948 and the frequency of the observed patterns for the 5 years period is given by the codes [1,X]. The digit "1" means the school is observed in that specific year, and the code "X" means the school is not observed. **(ii)** The positions of the codes [1,X] refer to the years along the panel, where the first position correspond to year 2005 and the fifth position (last) to year 2009.

The most common patterns in the original school panel are shown in Table 6.4. The first column indicates the combination year of school appearance in the panel given by the code 1 (2005, 2006,..., 2009), and the second column shows the number of observed schools for every spell. The first digit of the [1,X] combination code represents the first year of the panel (2005), and all subsequent digits up to the fifth represent the rest of years until 2009.

From the original selected samples used to estimate TEs and SEs between 2005 and 2009, we want to study the stability and trajectory of schools in terms of quality.[15] In order to do this, we need to construct a panel with trackable

[15]The descriptive statistics of all schools observed are shown in Table 8.29, Appendix 6.4.

schools along this period, from which we can analyse whether the schools experience changes in their school quality classification (*High, Mid-High, Mid-Low* and *Low*), focusing on those which improve or get worse over the period. Thus, in order to identify an upward or downward trajectory we require at least three observations per school.

We use only schools which are observed in 2005 and 2006 as baseline, and then we separate them into two main groups: (**Group 1**) schools for which their last SE estimation is observed in 2009; (**Group 2**) schools which we miss from the school panel in either 2008 or 2009. Within both groups we also retain schools which are not in the 2007 sample because we are aware of a specific problem with the original dataset for this year.[16] We then construct the RSP with a selection of trackable schools for at least three years, including 2005 and 2006.

As we observe in Table 6.5, around 76% of the RSP have at most one year of missing information regarding their the school quality level. We separate Group 2 from Group 1 because we may identify schools which have shifted to SS teacher scheme from 2008 or 2009 onwards.[17] Table 6.6 shows how many schools from the trackable groups shifted to the SS teacher scheme, representing in total 17% of the reduced school sample.

Table 6.5: Observed patterns in the reduce school panel

	Pattern	Fequency	Percent
Group 1	11111	296	38%
	11X11	126	16%
Group 2	1111X	167	22%
	111XX	97	13%
	11X1X	89	11%
Total		**775**	**100%**

Notes: (i) The reduced school panel (RSP) has 775 and the frequency of the observed patterns for the 5 years period is given by the codes [1,X]. The digit "1" means the school is observed in that specific year, and the code "X" means the school is not observed. **(ii)** The positions of the codes [1,X] refer to the years along the panel, where the first position correspond to year 2005 and the fifth position (last) to year 2009. **(iii) Group 1** is composed of schools who are always observed from 2005 to 2009, besides those schools which only disappeared in 2007. **(iv) Group 2** correspond to those schools from the RSP which we do not observe in 2009 or 2008 and 2009, considering also those missed in 2007.

[16] See in row (10), Table 6.1 for the selection sample at student level. The amount of observations dropped due to missing observations was almost twice larger compared to 2005 and three times compared to 2006.

[17] We identify schools to follow a subject specialist (SS) teacher scheme, when there are no more than one general teacher per grade within the school. From the selection sample description we observe an increasing pattern of schools shifting from general teachers to SS teachers. See in row (3), Table 6.1 how the percentage of observations eliminated increases because more SS teachers are teaching 4^{th} grade classes.

Table 6.6: Number of schools shifting to subject specialist (SS) teacher scheme

	Number of schools	Shifted to subject specialist (SS) teacher scheme			% shifted to SS
		2008	2009	Total	
	(1)	(2)	(3)	(4)	(5)
Group 1					
Group 1.1 (11111)	**296**	0	0	**0**	**0**
Group 1.2 (11X11)	**126**	0	0	**0**	**0**
Group 2					
Group 2.1 (1111X)	**167**	0	59	**59**	**35%**
Group 2.2 (11X1X)	**89**	0	27	**27**	**30%**
Group 2.3 (111XX)	**97**	41	6	**47**	**48%**
	775	41	92	**133**	**17%**

Notes: **(i)** Group 1 is composed of schools who are always observed from 2005 to 2009, besides those schools which only disappeared in 2007. **(ii) Group 2** correspond to those schools from the RSP which we do not observe in 2009 or 2008 and 2009, considering also those missed in 2007. **(iii)** Schools from **Group 2** could be missed from the panel in years 2008 or 2009 because they could have switched to the subject specialist (SS) teacher scheme and in columns (2) and (3) we count them. In column (5) we show the percentage of schools which switched to SS teacher scheme with respect to its group.

We compare the main characteristics of both school panels: the Original School Panel and the Reduce School Panel (RSP) in Table 6.7. In terms of average student performance along the period where schools are observed, there is a small increase in the standardised Simce score for the reduced panel with respect to the original panel. The distribution of type of school conditions remain similar between the two groups. The number of classes (equivalent to teachers), as well as the number of students per grade is slightly higher in the reduced school panel. The distributions of SEs and TEs have marginally moved to the right for the trackable schools.

The second part of Table 6.7, shows the school average of teacher characteristics. Here we do not observe considerable differences between the two panels. This is also true for principals characteristics, which for both panels are very close. The descriptive statistics support the idea that our analysis of SEs, and school quality stability may be representative of an important part of schools within the Chilean school system.[18]

[18] We have shown in Appendix 6.1 how similar are the original 4^{th} grade cohorts are compared to the selected sample due to estimation requirements.

Table 6.7: Descriptive statistics: Original vs Reduced School Panel
4^{th} grade cohorts: 2005 - 2009

	Original School Panel				Reduced School Panel			
Observed variables	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
School characteristics								
Average Stdz Languge Simce Score (4th)	0.00	0.44	-1.15	1.27	0.03	0.42	-1.11	1.12
Average Stdz Maths Simce Score (4th)	0.02	0.48	-1.31	1.36	0.05	0.46	-1.18	1.28
Municipal Schools	0.54	0.50	0	1	0.63	0.48	0	1
Private Voucher Schools	0.40	0.49	0	1	0.31	0.46	0	1
Unsubsidised Private Schools	0.06	0.24	0	1	0.06	0.23	0	1
Rural Area	0.04	0.19	0	1	0.03	0.16	0	1
Number of teachers (classes) per grade	2.2	0.5	2	5	2.4	0.6	2	5
Number of students per grade	62.2	21.5	30	203	70.6	24.1	35	203
School Quality (SE quartiles)	2.45	0.95	1	4	2.56	0.86	1	4
EB Language School Effects (SEs)	-0.01	0.25	-0.96	0.78	0.02	0.22	-0.69	0.66
EB Maths School Effects (SEs)	-0.01	0.22	-0.83	0.70	0.01	0.20	-0.62	0.59
Average EB Language Teacher Effects (TEs)	-0.01	0.13	-0.51	0.53	0.00	0.11	-0.39	0.33
Average EB Maths Teacher Effects (TEs)	-0.01	0.20	-0.76	0.78	0.01	0.17	-0.59	0.53
Teacher characteristics (school average)								
Gender (Female=1)	0.89	0.18	0	1	0.90	0.14	0.1	1
Age	47.5	7.6	24.0	79.5	48.5	6.4	31.0	64.4
Years of experience in the system	19.2	8.5	0.0	39.0	20.7	7.5	2.3	36.4
(Teaching hrs / Contract hrs) Ratio	0.89	0.12	0.2	1.0	0.89	0.10	0.2	1.0
Additional qualitfications	0.43	0.30	0.0	1.0	0.44	0.26	0.0	1.0
Post-graduate studies	0.03	0.09	0.0	1.0	0.02	0.07	0.0	0.8
Expect students complete technical degree	0.18	0.22	0.0	1.0	0.20	0.18	0.0	0.8
Expect students complete university degree	0.21	0.29	0.0	1.0	0.23	0.27	0.0	1.0
Cohorts specialist	0.67	0.29	0.00	1.00	0.71	0.23	0.00	1.00
Grade specialist	0.13	0.18	0.00	1.00	0.12	0.14	0.00	0.83
Newcomers	0.02	0.07	0.00	1.00	0.02	0.04	0.00	0.30
Other specialisation	0.20	0.23	0.00	1.00	0.16	0.15	0.00	1.00
Low quality Teachers	0.26	0.25	0.00	1.00	0.24	0.20	0.00	1.00
Mid-Low quality Teachers	0.25	0.22	0.00	1.00	0.25	0.16	0.00	0.83
Mid-High quality Teachers	0.25	0.21	0.00	1.00	0.25	0.16	0.00	0.83
High quality Teachers	0.24	0.24	0.00	1.00	0.26	0.21	0.00	1.00
Principal characteristics								
Gender (Female=1)	0.5	0.5	0	1	0.48	0.43	0	1
Age	58.5	9.8	29	85	58.7	8.7	34.3	85.0
Years of experience in the system	29.8	9.7	0	65.6	30.8	9.2	0.0	65.6
(Teaching hrs / Contract hrs) Ratio	0.4	0.4	0.00	1	0.43	0.42	0.0	1.0
Others								
Number of years observed	2.9	1.4	1	5	4.14	0.78	3	5
Without teacher variables (staff data base)	0.00	0.00	0	0	0.0	0.0	0.0	0.0
Without teacher variables (questionnaire)	0.03	0.09	0	1	0.0	0.1	0.0	0.5
Without principal variables (staff data base)	0.07	0.23	0	1	0.06	0.19	0.0	1.0
Number of students	370,519				230,911			
Number of teachers*	12,961				7,859			
Number of schools	1,948				775			

Notes: **(i)** The descriptive statistics of each variable correspond the average across schools given the average observed within school in 4^{th} grade cohorts over the period 2005 - 2009. **(ii)** The dummy variables show the proportion of each category in both school panels (Municipal, Private Voucher, Unsubsidised Private schools; Rural Area; Teacher's gender; Additional teacher qualifications; Post-graduate studies of teachers; Expectations on student completion; Cohort specialist; Grade specialist; Newcomers; Other specialisation; Low, Mid-Low, Mid-High, High quality teachers; Principal's gender) **(iii)** *The number of teachers counted here does not necessary mean unique teachers as there are teachers observed more than once along the period.

6.6.3 Defining school quality trajectories

The objective of this section consists of defining groups of schools by quality trajectories based on the yearly SE rankings. Here we consider the classification of school quality by quartiles from the original school panel, but we compare those changes among the group of trackable schools (or reduced school panel).

We construct four transition matrices from the base year 2005 to the destination categories in years 2006-2009. In every matrix we can observe whether a school from a specific quality category stays in the same category (along the diagonal), or moves to a lower (*downgrade*) or a higher category (*upgrade*). Downgrade or upgrade means to be observed in the lower triangle or upper triangle of the matrix, respectively. In the last two years (matrices 2008-2009), we add two destination categories in order to identify schools shifting to SS teacher scheme.

Table 6.8 shows movements between school quality categories with respect to the original classification in 2005. The three types of transition play an important role in defining the school quality trajectory. The *downgrade* movement is shaded with light-red colour, the *upgrade* transition is highlighted with light-blue colour while the stable categories are painted in light-grey. Notice that just for trajectory classification purposes, we consider those schools which keep *Low* quality category as downward movement, and the opposite to those staying in the *High* category which we include into the upward trajectory group. Following our broader classification of school category levels, we identify those schools which were not able to improve their effectiveness and treat them similarly than those which were getting worse over the period, as they cannot get worse either. Analogously, for those schools which stay in the highest category, we include them in the upward trajectory as there is not a higher category than *High*.[19]

Table 6.8: Transition matrices

		2006				2007				2008					2009				
		Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
2005	Low	79	53	26	11	59	33	22	5	68	41	21	13	8	39	27	12	4	19
	Mid-Low	49	61	32	32	40	41	33	13	43	45	48	18	11	30	21	26	17	20
	Mid-High	38	44	71	63	25	45	47	42	40	37	62	46	12	18	28	41	27	27
	High	12	29	59	116	6	23	55	71	8	44	49	95	10	8	25	35	64	26
		178	187	188	222	130	142	157	131	159	167	180	172	41	95	101	114	112	92
		Total 2005-06		**775**		**Total 2007**		**560**		**Total 2008**		**678**			**Total 2009**		**422**		

Notes: **(i)** The total number of schools in the reduced school panel (RSP) is 775, all of them are observed in 2005 and 2006. Schools are observed at least three times, but their last observations must be either in 2008 or 2009. **(ii)** In 2008 and 2009 we identify those schools who left the RSP because they shifted to a subject specialist (SS) teacher scheme, then we can not estimate TEs and SEs for them. **(iii)** All transitions are with respect to the base year 2005; the lower triangle (red-light colour) represents downward transitions, which we also include Low-Low combination as it cannot move to a lower category; the upper triangle (blue-light colour) shows upward movements, and we also consider the case High-High as there is not a higher quality level to move.

[19] See Appendix 6.5, the transition matrix tables disaggregated for every type of trackable group.

In percentage terms, we see Table 6.9 that from 2006 onwards, over the 52% of *High* quality schools were classified as *High* quality in the baseline year. For those schools which stay in the reduced panel until 2009, we observe that the proportion of schools staying in the higher quality category reaches the 57% of the total schools in this group. On the other hand, *Low* quality schools seems to be less stable than the highest ones. Only 41% from the lowest category in 2009 were in the same category in the baseline year, and their main transition is to *Mid-Low* category with a 27% of the schools.

Table 6.9: Transition matrices (percentage)

		2006				2007				2008					2009				
		Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
2005	Low	44%	28%	14%	5%	45%	23%	14%	4%	43%	25%	12%	8%	20%	41%	27%	11%	4%	21%
	Mid-Low	28%	33%	17%	14%	31%	29%	21%	10%	27%	27%	27%	10%	27%	32%	21%	23%	15%	22%
	Mid-High	21%	24%	38%	28%	19%	32%	30%	32%	25%	22%	34%	27%	29%	19%	28%	36%	24%	29%
	High	7%	16%	31%	52%	5%	16%	35%	54%	5%	26%	27%	55%	24%	8%	25%	31%	57%	28%

Notes: **(i)** The total number of schools in the reduced school panel (RSP) is 775, all of them are observed in 2005 and 2006. Schools are observed at least three times, but their last observations must be either in 2008 or 2009. **(ii)** In 2008 and 2009 we identify those schools who left the RSP because they shifted to a subject specialist (SS) teacher scheme, then we can not estimate TEs and SEs for them. **(iii)** All transitions are with respect to the base year 2005; the lower triangle (red-light colour) represents downward transitions, which we also include Low-Low combination as it cannot move to a lower category; the upper triangle (blue-light colour) shows upward movements, and we also consider the case High-High as there is not a higher quality level to move.

Depending on the observed transitions for every school, we create trajectory groups which suggest whether a school is improving its effectiveness, showing a positive trajectory, or on the contrary deteriorating with a negative movement. Therefore, we define two main groups of schools: (1) **Upward trajectory** schools, those which present enough evidence of a positive trajectory in their SE ranking; and (2) **Downward trajectory** schools, from which we observe the negative movement. In the following tables we show the set of conditions required for every trackable group in order to classify schools into any of the trajectory groups.

The first condition refers to the comparison between the school quality category in the baseline year and the last year in which the school is observed in the reduced panel. The second condition checks the transitions between the initial category level and the final one, and depending on the amount of years a particular school is observed, we look for a minimum of downward or upward evidence.

Particularly, for trackable groups 1.1 and 1.2 the first condition to be classified as a **Downward trajectory** school requires that the quality level of the school at the final year 2009 is lower than the quality level observed in the initial year 2005. We also include the case where the school stays in the same lowest category in both years. The second condition requires that for at least half of the cases where there could be a transition in quality category, the movements are stable or downward. Groups 2.1, 2.2 and 2.3 change the final year of condition 1, and the number of potential quality transitions to check in order to hold condition 2.

Table 6.10: Set of conditions - Downward trajectory schools

	Condition 1	Condition 2
Group 1.1	Quality Level 2005 > Quality Level 2009	at least 2 transitions (out of 3)
Pattern (11111)	or Quality Level 2005 = Quality Level 2009 = *Low*	Quality Level 2005 ≥ Quality Level 2006, 2007, 2008
Group 1.2	Quality Level 2005 > Quality Level 2009	at least 1 transitions (out of 2)
Pattern (11X11)	or Quality Level 2005 = Quality Level 2009 = *Low*	Quality Level 2005 ≥ Quality Level 2006, 2008
Group 2.1	Quality Level 2005 > Quality Level 2008	at least 1 transitions (out of 2)
Pattern (1111X)	or Quality Level 2005 = Quality Level 2008 = *Low*	Quality Level 2005 ≥ Quality Level 2006, 2007
Group 2.2	Quality Level 2005 > Quality Level 2008	Quality Level 2005 ≥ Quality Level 2006
Pattern (11X1X)	or Quality Level 2005 = Quality Level 2008 = *Low*	
Group 2.3	Quality Level 2005 > Quality Level 2007	Quality Level 2005 ≥ Quality Level 2006
Pattern (111XX)	or Quality Level 2005 = Quality Level 2007 = *Low*	

Notes: **(i)** The positions of the codes [1,X] refer to the years along the panel, where the first position correspond to year 2005 and the fifth position (last) to year 2009. **(ii)** Categorical variables Quality Level 2005 - 2009 indicate the school quality category in the respective year: (1) Low; (2) Mid-Low; (3) Mid-High; (4) High. **(iii)** For each group Conditions 1 and 2 have to hold. **(iv)** Condition 1 is a necessary but not sufficient condition, as Condition 2 is also required once Condition 2 happens. **(v)** Note that Condition 2 varies depending on the group (and number of observations available per school).

Table 6.11: Set of conditions - Upward trajectory schools

	Condition 1	Condition 2
Group 1.1	Quality Level 2005 < Quality Level 2009	at least 2 transitions (out of 3)
Pattern (11111)	or Quality Level 2005 = Quality Level 2009 = High	Quality Level 2005 ≤ Quality Level 2006, 2007, 2008
Group 1.2	Quality Level 2005 < Quality Level 2009	at least 1 transitions (out of 2)
Pattern (11X11)	or Quality Level 2005 = Quality Level 2009 = *High*	Quality Level 2005 ≤ Quality Level 2006, 2008
Group 2.1	Quality Level 2005 < Quality Level 2008	at least 1 transitions (out of 2)
Pattern (1111X)	or Quality Level 2005 = Quality Level 2008 = *High*	Quality Level 2005 ≤ Quality Level 2006, 2007
Group 2.2	Quality Level 2005 < Quality Level 2008	Quality Level 2005 ≤ Quality Level 2006
Pattern (11X1X)	or Quality Level 2005 = Quality Level 2008 = *High*	
Group 2.3	Quality Level 2005 < Quality Level 2007	Quality Level 2005 ≤ Quality Level 2006
Pattern (111XX)	or Quality Level 2005 = Quality Level 2007 = *High*	

Notes: **(i)** The positions of the codes [1,X] refer to the years along the panel, where the first position correspond to year 2005 and the fifth position (last) to year 2009. **(ii)** Categorical variables Quality Level 2005 - 2009 indicate the school quality category in the respective year: (1) Low; (2) Mid-Low; (3) Mid-High; (4) High. **(iii)** For each group Conditions 1 and 2 have to hold. **(iv)** Condition 1 is a necessary but not sufficient condition, as Condition 2 is also required once Condition 2 happens. **(v)** Note that Condition 2 varies depending on the group (and number of observations available per school).

Analogously in Table 6.11, we observe how the first necessary but not sufficient condition, requires for an improvement in the school quality level in the final year with respect to the baseline year. We include being stable in the highest group as a kind of good performance measure of school in terms of effectiveness. The additional second conditions also require that periods in between suggest a noticeable upward quality movement.

We apply the set of conditions presented in Table 6.10 and 6.11 to all the trackable groups belonging to the **reduced school panel**, and we get the following distribution of school by type of quality movement.

Table 6.12: Distribution of trajectories by trackable groups

	Reduced School Panel distibution			
	Downtrend	Other	Uptrend	Total
Group 1.1	108	85	103	**296**
Pattern (11111)	*36%*	*29%*	*35%*	
Group 1.2	60	20	46	**126**
Pattern (11X11)	*48%*	*16%*	*37%*	
Group 2.1	70	26	71	**167**
Pattern (1111X)	*42%*	*16%*	*43%*	
Group 2.2	40	16	33	**89**
Pattern (11X1X)	*45%*	*18%*	*37%*	
Group 2.3	47	16	34	**97**
Pattern (111XX)	*48%*	*16%*	*35%*	
Total	**325**	**163**	**287**	**775**
	42%	**21%**	**37%**	

Notes: **(i)** Columns of the matrix represent the type of school trajectory categories. **(ii)** Then, the matrix shows the distribution of every trajectory category by type of pattern. **(iii)** Each group has tow rows: the first row shows the number of schools, and the second row the percentage of the observation with respect to the total number of schools observed pattern group.

For those schools which do not show a clear pattern of downward or upward quality trajectory, we aggregate them into the **Other** category. The distribution of the three quality trajectories is relatively similar among the trackable groups (from 1.1 to 2.3). Only for the groups 1.1 and 2.1 with almost all possible school observations available, we see that proportions of **downward trajectory** and **upward trajectory** schools are very similar. In total, **upward trajectory** represents the 37% of schools in the reduced sample while **downward trajectory** group represents the 42%.

Table 6.13: Changes to subject specialist (SS) teacher scheme by trajectory groups

	Number of schools	Shifted to subject specialist (SS) teacher scheeme			% of the total group
		2008	2009	Total	
	(1)	(2)	(3)	(4)	(5)
Downtrend	**325**	17	39	**56**	**17%**
Other	**163**	6	16	**22**	**13%**
Uptrend	**287**	18	37	**55**	**19%**
	775	41	92	**133**	**17%**

Notes: **(i)** In 2008 and 2009 we identify those schools who left the RSP because they shifted to a subject specialist (SS) teacher scheme, then we can not estimate TEs and SEs for them. **(ii)** Column (4) is the total number of schools shifting to a subject specialist (SS) teacher scheme between 2008 and 2009 for each type of quality trajectory group; Column (5) correspond to the percentage of schools which switched to SS teacher scheme with respect to the total number of schools in the quality trajectory group.

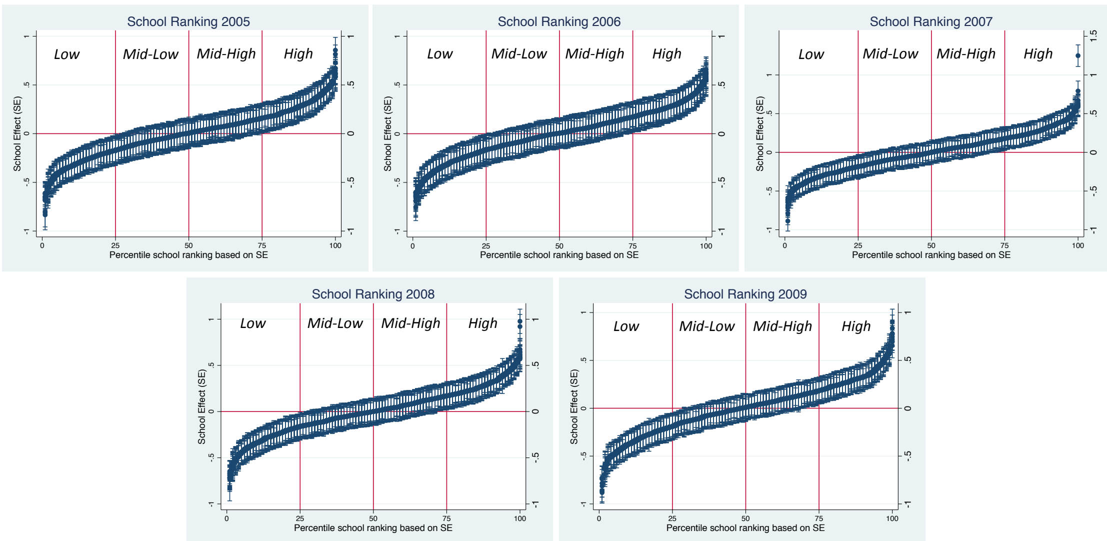
Regarding the schools which switched to the SS teacher scheme, we do not observe a significant difference between **downward trajectory** and **upward trajectory** schools, although the proportion is higher in the latest one. Therefore, the change into the SS teacher scheme seems not to be triggered by a particular downward or upward quality movement.

6.6.4 Static school ranking versus school quality movement analysis

Earlier in this chapter, we discussed some concerns related to school rankings and the use of LTs based on SE estimates. The uncertainty of SE predictions and the difficulty to rank schools precisely is one of the main issues to be addressed (Foley and Goldstein (2013)). Secondly, the stability of estimates over time, and whether schools maintain their quality level independent of the student cohort, are part of the discussion (McCaffrey et al. (2009); Goldstein (2014)).

To address the uncertainty of school rankings it is useful to group schools into broader quality categories, such as we have suggested using SE quartiles. Positions in school rankings might also vary over time, but these jumps might not be informative unless the differences between SE estimates are statistically significant.

Figure 6.3: School rankings 2005 - 2009



Notes: **(i)** School rankings based on estimated School Effects (SE), where the lowest percentile (1st) is for the 1st percent lower SE schools, and the 100th percentile corresponds to the 1 percent highest estimated SE schools. **(ii)** The percentile ranking is separated in quartiles, which we call the school quality categories: Low, Mid-Low, Mid-High and High. **(iii)** Whiskers represent the standard errors of SE estimates.

In Figure 6.3, schools are sorted from less to more effective in relation to their predicted school effects. Separating the percentile ranking in quartiles, we identify the four quality categories mentioned previously (*Low, Mid-Low, Mid-High* and *High*). Similar to what Leckie and Goldstein (2011) have found, we observe that SE estimates are significantly different from zero in the lowest and highest quartile, while most of estimates in the middle are not statistically different from zero. The standard errors, represented by SE whiskers, overlap each other in closely ranked schools, suggesting that SE estimates are not significantly different from each other among them. This fact confirms the uncertainty issues regarding school ranking predictions.

Static school rankings would lead us to similar uncertainty concerns ever year, while if we focus on movements between broader school quality groups, and particularly in their trajectories, we minimise the ranking misclassification problems. It is important to highlight, that even using the school quality categories, we still might be misclassifying some schools. However, taking advantage of the available data, we are able to observe from three to five SE estimates per school in the RSP. Thus, we can have a better idea how schools are progressing in terms of effectiveness along the period of study.[20]

We suggest that analysing school quality trajectories would ameliorate school misclassification problems. In **downward trajectory** schools we do not only identify schools which have decreased their quality level, but we also identify schools that have been mostly classified as *Low*. Analogously in **upward trajectory** schools, where we observe schools which have been constantly effective and those which have improved during the period. The classification of both types of quality trajectories, **downward trajectory** and **upward trajectory**, is useful to understand the SE estimates in the long run. This classification might be helpful as well, to implement focused policies, and assess their results for reward or intervention purposes.

In the following section we try to explain what determines whether a school has a **downward trajectory** or **upward trajectory** in its quality for at least a three year period.

6.7 Determinants of school quality trajectories

The main objective of this paper is to find determinants of school effectiveness improvements, which at first sight seems not be driven either by school or teacher observable characteristics. Among trackable schools we have found that the proportion of **downward trajectory** schools is not noticeably larger than **upward trajectory** schools.

To compare these groups of schools, we show in Table 6.14 the descriptive statistics in terms of school, principal and teacher characteristics at school level. It is clear that the averages of standardised Simce scores (Language and Maths) are the result of schools effectiveness (SE) along the period. However, we attempt to explain the significant differences between the three groups.

[20]If we only focus in yearly SE estimations, without analysing their effectiveness evolution, we could also arrive to some misleading conclusions. See Figure 6.3, Appendix 6.6, the static school rankings by type of quality trajectory. If we analyse these plots separately, they are not very informative on how well or badly schools are doing.

Table 6.14: Descriptive statistics: Type of quality trajectory
4^{th} grade cohorts: 2005 - 2009

	Downtrend				Uptrend				Other			
Observed variables	**Mean**	**Std.Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
School characteristics												
Average Stdz Languge Simce Score (4th)	-0.09	0.42	-1.11	1.12	0.15	0.39	-0.61	1.122	0.15	0.39	-0.61	1.12
Average Stdz Maths Simce Score (4th)	-0.07	0.46	-1.18	1.28	0.16	0.43	-0.64	1.28	0.16	0.43	-0.64	1.28
Municipal Schools	0.63	0.48	0	1	0.62	0.49	0	1	0.62	0.49	0	1
Private Voucher Schools	0.31	0.46	0	1	0.33	0.47	0	1	0.33	0.47	0	1
Unsubsidised Private Schools	0.06	0.23	0	1	0.05	0.21	0	1	0.05	0.21	0	1
Rural Area	0.03	0.18	0	1	0.02	0.13	0	1	0.02	0.13	0	1
Number of teachers (classes) per grade	2.4	0.6	2	5	2.4	0.6	2	5	2.4	0.6	2	5
Number of students per grade	70.2	24.2	35.33	202.8	71.0	24.5	34.75	179.8	71.0	24.5	34.75	179.8
School Quality (SE quartiles)	2.23	0.84	1	3.8	2.89	0.85	1.2	4	2.89	0.85	1.2	4
EB Language School Effects (SE)	-0.07	0.22	-0.69	0.45	0.11	0.23	-0.38	0.66	0.11	0.23	-0.38	0.66
EB Maths School Effects (SE)	-0.06	0.19	-0.62	0.40	0.10	0.20	-0.34	0.59	0.10	0.20	-0.34	0.59
Average EB Language Teacher Effects (TE)	-0.02	0.11	-0.39	0.29	0.03	0.11	-0.28	0.33	0.03	0.11	-0.28	0.33
Average EB Maths Teacher Effects (TE)	-0.03	0.18	-0.59	0.46	0.05	0.17	-0.43	0.53	0.05	0.17	-0.43	0.53
Teacher characteristics (school average)												
Gender (Female=1)	0.88	0.15	0.1	1	0.91	0.14	0.125	1	0.91	0.14	0.125	1
Age	48.1	6.44	31.6	64.4	49.0	6.4	31.38	60.6	49.0	6.43	31.38	60.56
Years of experience in the system	19.9	7.51	2.3	34.4	21.4	7.5	3.5	36.4	21.4	7.46	3.5	36.42
(Teaching hrs / Contract hrs) Ratio	0.90	0.10	0.33	1	0.89	0.11	0.18	1	0.89	0.11	0.18	1
Additional qualitfications	0.44	0.26	0	1	0.45	0.26	0	1	0.45	0.26	0	1
Post-graduate studies	0.03	0.07	0	0.38	0.03	0.07	0	0.75	0.03	0.07	0	0.75
Expect students complete technical degree	0.20	0.18	0	0.83	0.20	0.18	0	0.75	0.20	0.18	0	0.75
Expect students complete university degree	0.19	0.26	0	1	0.27	0.28	0	1	0.27	0.28	0	1
Cohorts specialised	0.70	0.24	0	1	0.71	0.22	0	1	0.71	0.22	0	1
Grade specialised	0.12	0.14	0	0.83	0.13	0.16	0	0.83	0.13	0.16	0	0.83
Newcomers	0.02	0.04	0	0.2	0.02	0.05	0	0.27	0.02	0.05	0	0.27
Other training	0.17	0.17	0	1	0.15	0.14	0	0.63	0.15	0.14	0	0.63
Low quality Teacher	0.28	0.22	0	0.89	0.20	0.19	0	1	0.20	0.19	0	1
Mid-Low quality Teacher	0.27	0.17	0	0.83	0.23	0.16	0	0.67	0.23	0.16	0	0.67
Mid-High quality Teacher	0.23	0.15	0	0.75	0.25	0.16	0	0.72	0.25	0.16	0	0.72
High quality Teacher	0.22	0.19	0	0.92	0.31	0.23	0	1	0.31	0.23	0	1
Principal characteristics												
Gender (Female=1)	0.46	0.43	0	1	0.49	0.44	0	1	0.49	0.44	0	1
Age	58.8	8.8	34.33	85	58.9	8.8	38	85	58.9	8.8	38	85
Years of experience in the system	30.5	9.2	0	52.8	31.4	9.2	0.5	65.6	31.4	9.2	0.5	65.6
(Teaching hrs / Contract hrs) Ratio	0.42	0.41	0	1	0.44	0.42	0	1	0.44	0.42	0	1
Others												
Number of years observed	4.06	0.7731	3	5	4.13	0.7607	3	5	4.13	0.76	3	5
Without teacher variables (staff data base)	0	0	0	0	0	0	0	0	0.0	0.0	0	0
Without teacher variables (questionnaire)	0.02	0.05	0	0.33	0.02	0.04	0	0.25	0.0	0.0	0	0.25
Without principal variables (staff data base)	0.07	0.20	0	1	0.05	0.19	0	1	0.05	0.19	0	1
Number of students	94,419				85,527				50,965			
Number of teachers*	3,247				2,894				1,718			
Number of schools	325				287				163			

Notes: (i) The descriptive statistics of each variable correspond the average across schools given the average observed within school in 4^{th} grade cohorts over the period 2005 - 2009. **(ii)** The dummy variables show the proportion of each category in both school panels (Municipal, Private Voucher, Unsubsidised Private schools; Rural Area; Teacher's gender; Additional teacher qualifications; Post-graduate studies of teachers; Expectations on student completion; Cohort specialist; Grade specialist; Newcomers; Other specialisation; Low, Mid-Low, Mid-High, High quality teachers; Principal's gender) **(iii)** *The number of teachers counted here does not necessary mean unique teachers as there are teachers observed more than once along the period.

As we want to explain the determinants of the quality trajectories, we analyse the observed and unobserved characteristics of these groups of school. Starting from the general descriptive statistics in Table 6.14, we rule out any difference due to the type of school dependence. *Municipal*, *Private Voucher* and *Unsubsidised Private* schools are similarly distributed among quality trajectories. Neither rurality conditions, number of classes per grade nor number of student per grade vary among the trajectory groups. What actually seems to differ between **downward trajectory** and **upward trajectory** schools is the average of the

TEs, that we also use to define teacher quality levels.

Both TEs and SEs are assumed to be *iid*, and holding the set of assumptions proposed earlier from ***A1*** to ***A3***, they are uncorrelated with each other, with other covariates, and with the error term. The distributions of TEs and SEs show the impact of unobserved heterogeneities on pupil achievement at teacher and school level, respectively. Every year, each school has an estimated SE and an average TE estimate, and for the all observations within each quality trajectory group we obtain the average shown in Table 6.14.

Independent of the school quality level where the school is moving to in each trajectory group, we clearly observe that the average of TE estimates in Language 0.03 and Maths 0.05 are higher in **upward trajectory** schools compared to **downward trajectory** schools, with -0.02 and -0.03, respectively. There might be unobserved teacher skills that might boost the unobserved school factors along this period. Thus, it is interesting to analyse how the evolution of these groups of school is in terms on the teacher quality composition within them.

With respect to the teacher observed characteristics, they are collapsed yearly at school level, representing the average of the variables in each 4^{th} grade cohort. Thus, the average of female teachers in all types of trajectory schools is approximately 90%, their age is around 48 and their years of experience 20. Also the teaching ratio and additional qualifications looks very similar between groups, and only the expectations on students completing a university degree seems to be significantly higher in upward trajectory schools. The distribution of type of teacher specialisation is even among the groups, and we will confirm in the next subsection that type of teacher training seems not to determine school effects.

Regarding principal characteristics we do not find significant differences either among type of schools movement. Their gender, age, years of experience and teaching ratio is basically the same.

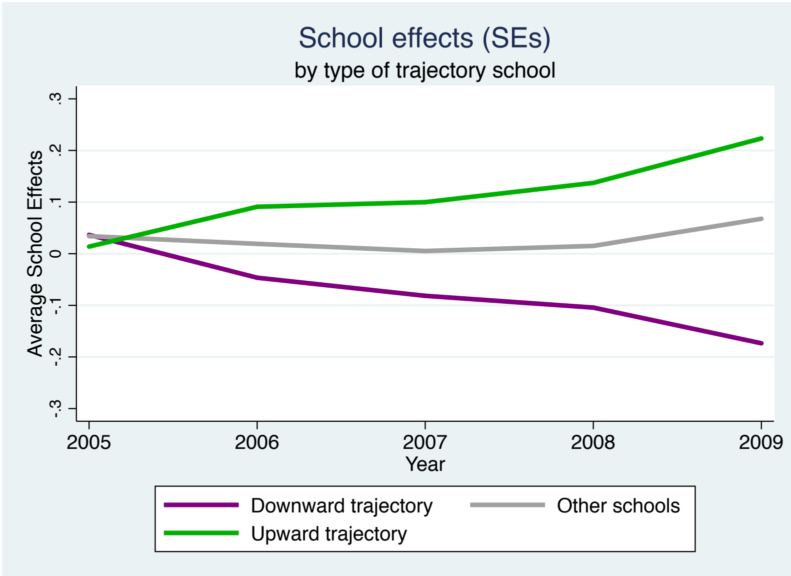
The next subsection presents graphical evidence of the descriptive statistics discussed above, putting special emphasis on the type of teacher specialisation and teacher quality levels per school.

6.7.1 School effects evolution by type of schools

In the previous subsection we constructed three trajectories of school quality levels based on SE effects rankings among the whole sample of the original school panel. Then, considering the reduced school panel (RSP) with only trackable schools we formed the three groups of school quality level stability: **downward trajectory**, **upward trajectory** and **other** as shown in Figure 6.4 below.

By definition, the school quality trajectories have to show a clear tendency in the estimated SEs (either upward or downward), as it is shown in Figure 6.4. In 2005 all schools in the RSP present similar averages of SE estimates, while the gap start increasing since 2006.

Figure 6.4: School effects evolution 2005 - 2009



Additionally, we found that part of the attrition from the school panel is due to schools switching to SS teacher scheme (17% approx.). However, from Table 6.13 in the previous section, we could not identify any particular pattern in the SE evolution that might explain the change from general teachers to SS teachers.

Moreover, we can observe in Figure 6.5 that schools from the **upward trajectory** group which switched to the SS teacher scheme in 2008 were above the rest of schools remaining in the panel in terms of SE estimates. Nevertheless, the opposite happens for the **downward trajectory** group, where those schools that shifted to the SS teacher scheme were below the average of SE with respect to the other schools which remain in the panel. Thus, shifting to the SS teacher scheme seems not to be triggered by a particular movement on the school quality trajectory.

In terms of type of school dependence, the school quality trajectories are not explained either by differences on *Municipal* or private schools (*Voucher* or *Unsubsidised*). Actually, we can infer from the distribution of quality trajectories by school dependence presented in in Figure 6.6, that *Unsubsidised Private* schools have significant lower proportion of **upward trajectory** schools compared to the rest of schools which rely on public funds.

Figure 6.5: School effects evolution 2005 - 2009
With and without subject specialist (SS) teacher scheme

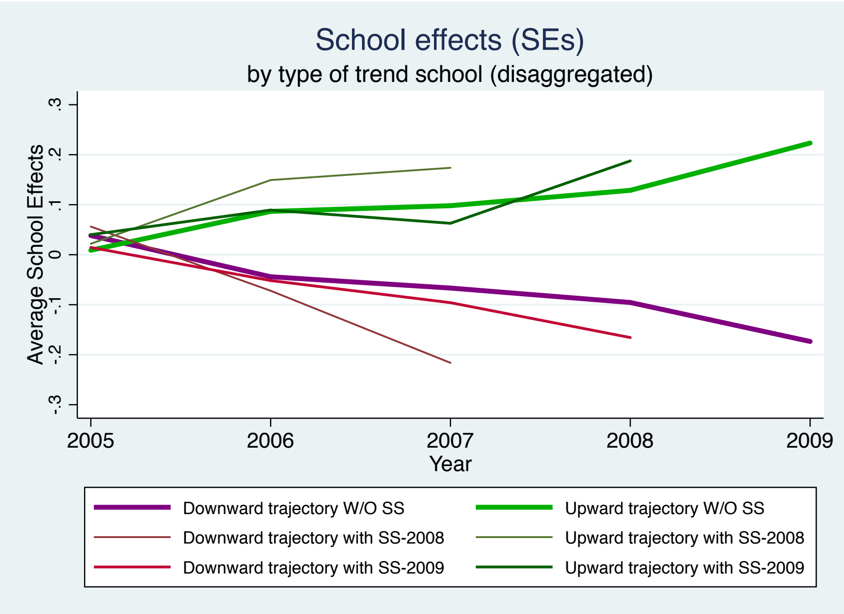
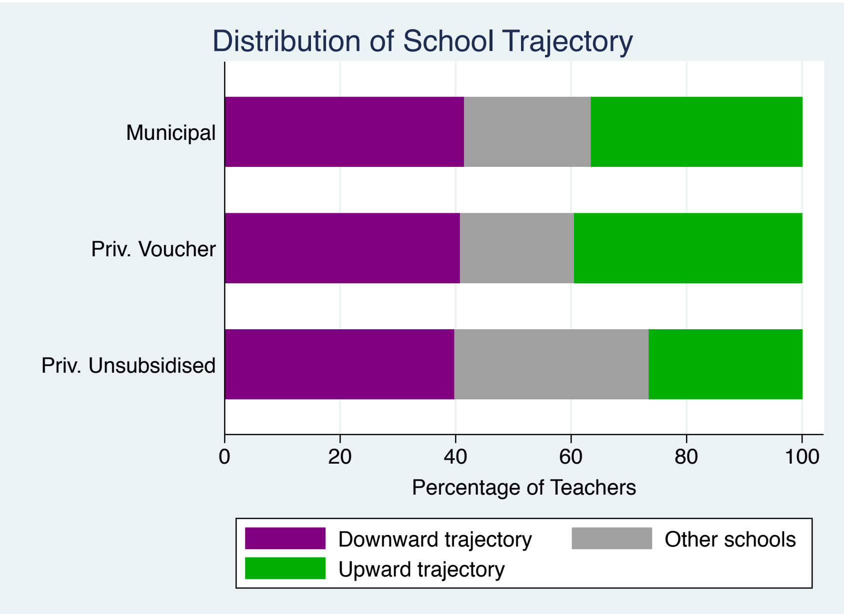


Figure 6.6: School trajectory 2005 - 2009
by type of school dependence



We confirm that the characteristics of school itself do not explain the evolution of SEs. On the contrary, it seems that the distribution of the type of schools by quality trajectory is relatively even. In the following subsection, we focus on whether teacher characteristics can explain differences in the school quality trajectories.

6.7.2 School effects evolution by type of teachers

From the initial sample selection, we have focused on schools with at least two classrooms, and which have been taught by general teachers. These conditions are necessary to estimate SEs and TEs simultaneously. All teacher characteristics in the school panel are collapsed at school level, where every observation represents yearly averages per school, such as the proportion of *Low*, *Mid*, and *High* quality teachers per school-grade-year.

The average of observable school and teacher characteristics seems to be equally distributed among school quality trajectories. As we have already observed in Table 6.13, the means of the type of school dependence; rural conditions; number of students per grade; the proportion of teacher's gender; teacher's age; teacher's experience; and additional teacher's qualification do not considerably differ between the type of SE trajectories.

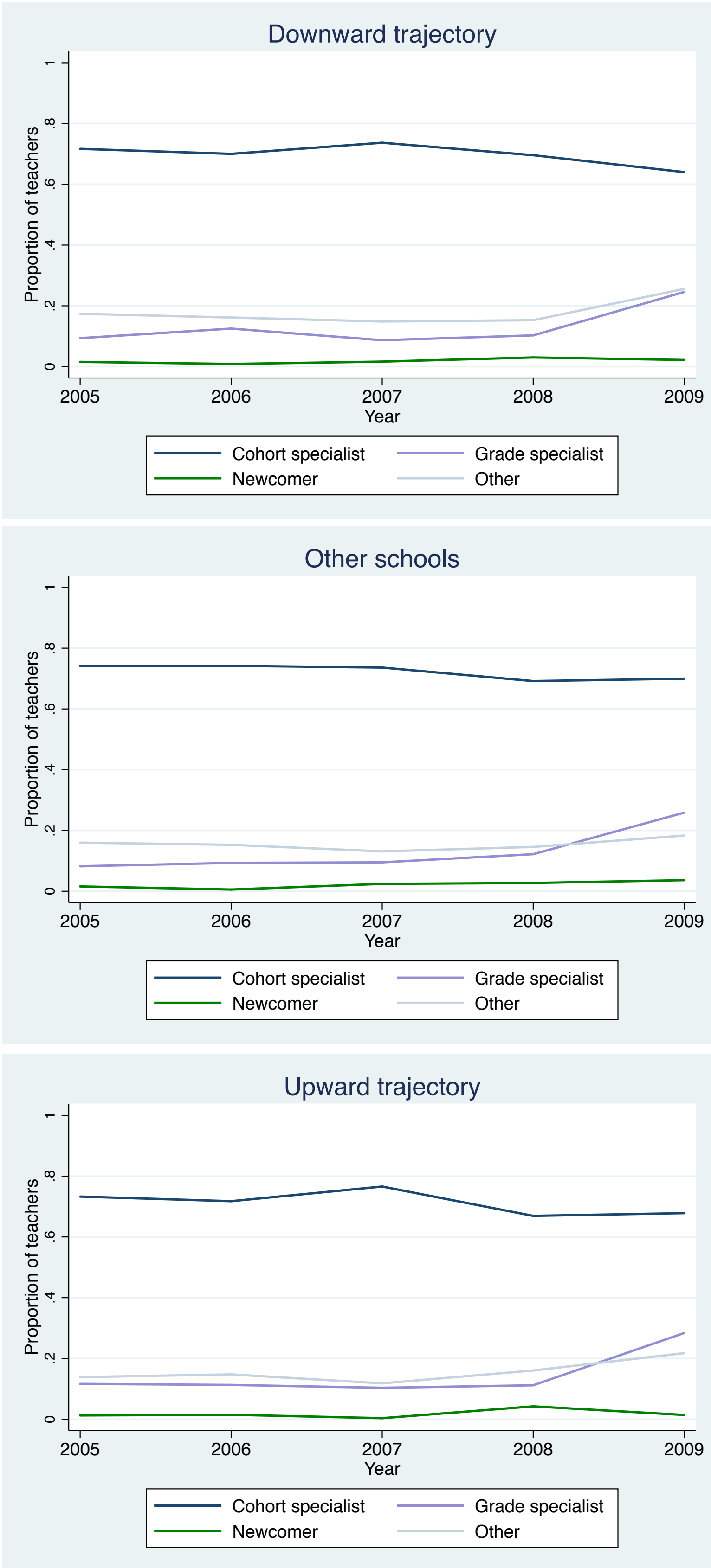
In addition, we have recovered the history of 4^{th} grade general teachers, allowing us to identify the type of teacher specialisation in previous grades and class allocation, before teaching students in 4^{th} grade cohorts between 2005 and 2009. As we have mentioned in the previous section, we classified them as: (i) **Cohort specialist** (CS), (ii) **Grade specialist** (GS), (iii) **Newcomers**, and (iv) **Others** teachers.[21]

Figure 6.7 suggests that the proportions of type of teacher specialisation do not vary considerably among types of school quality trajectory.[22] This evidence does not suggest any correlation between the type of teacher specialisation and the school effectiveness performance.

[21]It is important to highlight that we did not use this classification to estimate neither SEs nor TEs in the VAMs.

[22]See Figure 8.31, Appendix 6.3, the disaggregated teacher specialisation evolution. Here, it is clear how from 2007 the proportion of 4^{th} grade teachers teaching the same classroom-cohort from 2^{nd} grade decrease dramatically across all school quality trajectories. On the other hand, despite of the smaller variation, all types of SE trajectory have experienced an increased in the type of teacher specialised in 4^{th} grade for at least 2 years.

Figure 6.7: Teacher specialisation evolution 2005 - 2009
by school quality trajectory



The set of unobserved teacher heterogeneities represented by estimated TEs might be associated to the variation of SEs along this period (2005-2009). Thus, similar to the school quality category, we create a classification of teacher quality levels, given the yearly ranking of teachers based on estimated TEs. In the lowest quartile of the distribution, we classify teacher as *Low*, the second quartile the *Mid-Low*, the third quartile *Mid-High*, and teacher from the highest quartile as *High*.

The classification of teacher quality levels is created using the original selected cohort for each year, therefore the categories might not be equally distributed in the RSP. Thus, it is interesting to observe how *Low* and *High* quality teacher are distributed across schools by type of school quality trajectories.

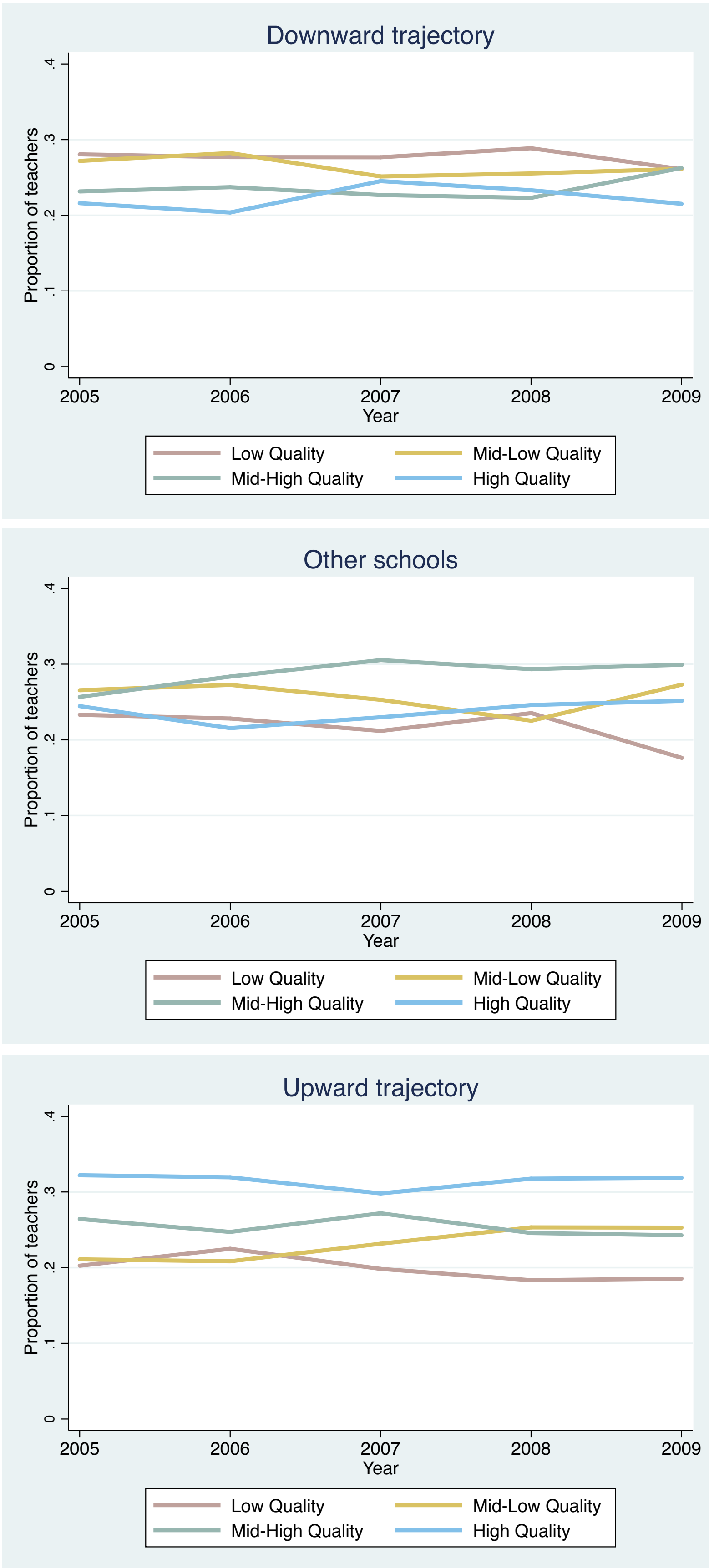
Comparing the means of estimated TEs within schools, we observe in Figure 6.8 that the proportion of teacher quality categories is relatively stable over the period in every type of school quality trajectory. Within **downward trajectory** schools, the *Low* quality teachers present the highest proportion in most of the years, with a very similar proportion of *Mid-Low* quality teachers in at least three out of five years. Schools without a defined trajectory, or **others**, are mainly composed of *Mid-High* quality teachers during this period.

Finally, for **upward trajectory** schools, there is a noticeable difference between the *Low* and *High* quality teacher proportions along the whole period. While the *High* quality group constantly represents over the 30% of teachers always, the proportion of *Low* quality is around 20%. Despite teacher quality proportions being relatively close in **downward trajectory** schools and **other** schools, the strictly dominance of *High* quality teachers in **upward trajectory** schools is distinguished. These findings suggest a potentially dynamic relationship with school effects might improving when their teacher effectiveness is high and stable.

We are aware that both TEs and SEs capture the variation in pupil's academic performance from unobserved heterogeneities at teacher and school level, respectively. Hence, the relationship between these two unobserved factors can be associated to the static TEs and SEs estimation. However, we do not observe anything in our estimation methodology that could explain such a clear and stable relationship between the estimated TEs in a particular year and the differences in the SEs estimates along the period.

The relationship between the levels of TEs and the differences in levels of SEs suggest that the composition of teachers within schools is an important factor to improve school quality in the long run.

Figure 6.8: Teacher quality evolution 2005 - 2009
by school quality trajectory



6.8 Conclusion

In this chapter we have discussed estimates of school effects (SEs) and in particular the issues associated with League Tables (LTs) based on them. We review the relevant literature, from the uncertainty of estimates to the potential negative consequences of publicly available school rankings. The difficulty of comparing schools one-to-one discourages the use of LTs to reward or punish institutions based on these rankings. Additionally, the responses of schools and users of the LTs might worsen student and teacher selection issues.

Nevertheless, we are aware of the advantages of using available information for statistical inference and consistent estimation of school quality. SE measures are useful for accountability purposes and for improving the allocation of resources at public or private level, depending on the schools and students needs.

Therefore, the objective of this chapter is to contribute to the SEs literature with a different perspective of how to assess and use SE estimates. We have proposed to analyse school effectiveness in the long run in order to minimise uncertainty issues associated with school rankings based on single measures. Moreover, we are interested in the evolution of school effectiveness and the factors that explain positive or negative trajectories.

Based on common VAM specifications and estimation strategies, we estimate SEs and TEs yearly from 2005 to 2009, where we were able to classify schools in four different school quality categories (*High, Mid-High, Mid-Low*, and *Low*). We then selected a group of trackable schools with at least three measures of SEs available. From this reduced school panel we were able to observe the evolution of school effectiveness in terms of their quality level classification.

We created two main groups of school representing improving or decreasing their school effectiveness in the long run. To do that, we define specific criteria for **upward** and **downward** school effectiveness movements along the period of study. Focusing on trajectories rather than just static measures helps to correct for potential schools quality misclassifications in schools close to the cutoffs.

From our analysis, we find that school effectiveness movements are not explained by observable characteristics of students, teachers and schools. Furthermore, the type of school dependency does not seem to explain how effective schools are in the long run. Taking advantage of teachers history track, we also ruled out that the type of teacher specialisation may determine the trajectories. What seems to drive **upward** (or **downward**) effectiveness movements is the proportion of high quality (or low quality) teachers per year-grade within schools.

These findings contribute to the teacher and school effectiveness literature in terms of the stability of single cohort measures for an specific period of time. The relationship observed between the composition of teacher quality within schools and the school quality trajectory, enhance the importance of teacher unobserved skills and the relevance of TE and SE estimations in potential educational programmes.

Chapter 7

Summary and Conclusions

The literature has explored estimations of Teacher Effectiveness (TEs) and School Effectiveness (SEs) in several educational contexts. These estimations are routinely obtained from Value Added Models (VAMs), which can be inferred from a general achievement function (GAF). In Chapter 2, we derived the most common VAMs, and discussed in detail two estimation approaches with typical specification restrictions and the respective assumptions required to consistently estimate TEs and SEs.

Teacher and school Value Added estimates are based on the respective capability of teachers and schools to improve individual *cognitive* abilities, which are represented by measures of pupil academic performance. The measures of individual achievement are usually obtained from standardised examination scores. In our case, we base our analyses on the Chilean educational context, and we use the standardised Chilean Examination (Simce), which is taken on a yearly basis and at national level, though not for all grades. The VAM specifications we have derived to estimate TEs and SEs require two consecutive individual achievement measures. Therefore, in Chapter 3, apart from presenting our data, we have shown that school marks standardised at school level are good proxies to substitute for standardised Simce scores, when the Simce scores are required but not available.

The consistency of teacher and school effects estimators observed in the literature rely on assumptions such as; random assignment of pupils to schools and classrooms, and random assignment of teachers to schools. These assumptions are difficult to test in the non-experimental frameworks that typify most of the educational contexts. However, some authors (e.g. Guarino et al. (2014b); Dieterle et al. (2015); Guarino et al. (2015)) have studied whether VAM estimators are able to predict true TEs under different scenarios using simulated data. They find non-significant differences between estimators predicting teacher quality rankings when there is random assignment of pupils to classrooms, although the ranking predictions might differ when there is evidence of non-random assignment. Additionally, Guarino et al. (2014a,b) conclude that the most robust TE estimators across all scenarios are related to the Empirical Bayes (EB) approach, under which estimated TEs predict successfully true teachers ranking in random assignment scenarios.

To better understand the Chilean educational context, in Chapter 4, we have checked for evidence of non-random assignment of students to teachers or classrooms in the Chilean primary schools. We have analysed three 4^{th} grade cohorts (2005, 2007, 2009), generating a mini-panel of three years, from 3^{rd} to 5^{th} grades, for each cohort. We follow graphical analyses, and conduct statistical tests in which we compare real class distributions with artificially created counterfactual class distributions, based on previous performance measures at the school-grade level. We present the analysis for the 4^{th} grade cohort, where we do not observe any graphical evidence of non-random assignment, and with respect to the statistical tests (**t-test**, **ks-test**) we find that 64% of schools do not show any evidence of non-random assignment of pupils to teachers or classrooms, and 93% of schools are classified with *No* or *Low* levels of Sorting Evidence (SoE) of pupils to teachers or classrooms (similar results are found for the other two 4^{th} grade cohorts).

We exploit our the uniquely rich data, and make use of the evidence of random assignment of pupils to classrooms in the Chilean primary schools to estimate our VAMs derived in Chapter 2. We use the Maximum Likelihood estimation (MLE) methodology, and we obtain predictions of teacher and school effects from the EB distributions. Hence, in Chapter 5 we have presented the first teacher and school Value- Added measures for the Chilean school system. We have found that 4^{th} grade general teachers have larger impacts on pupils Maths achievement than on Language achievement. For example, if a representative student from the 50^{th} percentile of the Simce score distribution is taught by a teacher who is one standard deviation more effective, that student is expected to ascend approximately 9 percentile ranking positions in Language, and 12 percentile ranking positions in Maths. Nevertheless, it is important to be aware that this TEs interpretation relies on very restrictive assumptions that limits its use as an instrument for policy implementation.

In Chapter 5, we also have checked for heterogeneous teacher effects. We analysed separately TE estimates from single-sex schools and compared them with those obtained from the whole cohort sample. The results do not seem to be very robust as the size of the sub-samples are very small. Additionally, we also disaggregate the original selected 4^{th} grade 2005 cohort into two sub-samples by type of school dependence. In the first sub-sample, we removed *Unsubsidised Private schools*, and in the second we considered only *Municipal schools*. We found that TEs are more heterogeneous in *Municipal schools* using both sub-samples. These finding imply two conclusions: (i) teachers from *Municipal schools* might have a larger impact (positive or negative) on pupil academic performance, (ii) heterogeneous teacher effects, with respect to the school dependence, can be equally estimated from different types of school samples.

In relation to the SEs, using single measures of estimations to construct school rankings or league tables (LTs) may lead to misclassification of school rankings due to large standard errors associated with the SEs. Hence, we have taken advantage of having Chilean education datasets for an extended period of time, and we analyse movements of school quality levels instead. Our analysis is focused on the evolution of SE estimates over a period of time, and we relate them to the types of teachers, in terms of quality and verified specialisation.

In Chapter 6, we have found that school effectiveness trajectories are not explained by observable characteristics of students, teachers or schools. Furthermore, the type of school dependency does not seem to be related with high-effective or low-effective school stability in the long run. Exploring teachers track records, we also ruled out the possibility that the type of teacher specialisation determines the trajectories. What seems to drive effectiveness movements **upward** (or **downward**) is the proportion of *high-quality* (or *low-quality*) teachers per year-grade within schools. These findings highlight the importance of teacher unobserved skills, and the relevance of TEs on school quality improvements.

The results shown in Chapters 4 and 5 have been obtained using two different VAM specifications: one with a preset value of the persistence parameter λ, and another with an unrestricted value of λ. In both cases, the results suggest that teachers are important in improving pupil academic performance. Additionally, assuming a preset value of the persistence parameter λ in **Chapters 5**,we have found that teacher quality level is associated with the stability of SE estimates, and larger proportions of *high-quality* teachers within schools are related to the improvements of school effectiveness in the long run.

Further research

Recently, there has been increasing interest in educational studies related to school and teacher quality. Simultaneously, more governments are allowing access to large and rich administrative datasets related to individual academic records, along with detailed information on pupils, families, teachers, and schools. The Chilean Government provides a prime example; its Ministry of Education has promoted the research in education, allowing the access to this type of data subject to a research proposal.

The analyses presented in this thesis make use of part of the huge amount of available data. Therefore, I plan to continue working with these data to extend my research agenda on students academic performance, by addressing topics such as: effects of principals, municipalities or local areas administration , and incentive payment schemes for teachers. We will study long-term impacts (as evidenced by students higher education performance) of teachers, principals, and schools.

In the short run, we will link our current research with principal effect estimations. From our available data, we are able to recover principals' characteristics, and we can track their movements and job positions across schools. Hence, we will study how changes of principal are related to school quality trajectories and the teacher composition within schools.

Additionally, we plan to investigate whether a principal's gender affects teacher and school effectiveness. Currently, the gender of principals is randomly distributed across schools, 50/50 approximately. We could take advantage of principal movement between schools to explore whether the random allocation of principal gender has an impact on SEs trajectories and teacher composition within schools. However, identifying effects of principal and teacher-gender matching is less feasible, as in early primary schools most teachers (90 percent) are female.

We could also estimate differences between general teacher effects and subject specialist (SS) teacher effects in particular subjects. The estimation of SS teacher effects, simultaneously with general teacher effects, is a computationally intensive undertaking that is already in progress.

Bibliography

Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in the chicago public high schools. *Journal of Labor Economics*, 25(1):95–135.

Andrabi, T., Das, J., Khwaja, A. I., and Zajonc, T. (2011). Do value-added estimates add value? accounting for learning dynamics. *American Economic Journal: Applied Economics*, pages 29–54.

Angrist, J. D. and Lavy, V. (1997). Using maimonides' rule to estimate the effect of class size on student achievement. Technical report, National Bureau of Economic Research.

Ballou, D., Sanders, W., and Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1):37–65.

Blanden, J. and Machin, S. (2010). Changes in inequality and intergenerational mobility in earlyyears assessments in hansen, k. *Joshi, H. Dex, S.(eds) Children of the 21st century*, 2.

Boardman, A. E. and Murnane, R. J. (1979). Using panel data to improve estimates of the determinants of educational achievement. *Sociology of education*, pages 113–121.

Buddin, R. (2011). Measuring teacher and school effectiveness at improving student achievement in los angeles elementary schools. Technical Report 31963, Munich Personal RePEc Archive.

Buddin, R. and Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2):103–115.

Carrington, B., Tymms, P., and Merrell, C. (2008). Role models, school improvement and the 'gender gap'—do men bring out the best in boys and women the best in girls? 1. *British Educational Research Journal*, 34(3):315–327.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4):778–820.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., and York, R. (1966). Equality of educational opportunity. *Washington, dc*, pages 1066–5684.

Cox, C. (2004). Innovation and reform to improve the quality of primary education: Chile. *Paper commissioned for the EFA Global Monitoring Report 2005, The Quality Imperative.*

Dearden, L., Machin, S., and Reed, H. (1997). Intergenerational mobility in britain. *The Economic Journal*, pages 47–66.

Dieterle, S., Guarino, C. M., Reckase, M. D., and Wooldridge, J. M. (2015). How do principals assign students to teachers? finding evidence in administrative data and the implications for value added. *Journal of Policy Analysis and Management*, 34(1):32–58.

Doris, A., O'Neill, D., and Sweetman, O. (2013). Gender, single-sex schooling and maths achievement. *Economics of Education Review*, 35:104–119.

Foley, B. and Goldstein, H. (2013). Measuring success: League tables in the public sector. British Academy.

Friedman, M. (1955). *The role of government in education.* Rutgers University Press.

Friedman, M. (1997). Public schools: Make them private. *Education Economics*, 5(3):341–344.

Gibbons, J. D. and Chakraborti, S. (2011). *Nonparametric statistical inference.* Springer.

Goldstein, H. (2011). *Multilevel statistical models*, volume 922. John Wiley & Sons.

Goldstein, H. (2014). Using league table rankings in public policy formation: Statistical issues. *Annual Review of Statistics and Its Application*, 1:385–399.

Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 385–443.

Guarino, C. M., Maxfield, M., Reckase, M. D., Thompson, P., and Wooldridge, J. M. (2014a). An evaluation of empirical bayes' estimation of value-added teacher performance measures. Technical Report 31, Education Policy Center at Michigan State University Working Paper.

Guarino, C. M., Reckase, M. D., Stacy, B. W., and Wooldridge, J. M. (2015). Evaluating specification tests in the context of value-added estimation. *Journal of Research on Educational Effectiveness*, 8(1):35–59.

Guarino, C. M., Reckase, M. D., and Wooldridge, J. M. (2014b). Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, 10(1):117–156.

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, pages 280–288.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of economic literature*, pages 1141–1177.

Hanushek, E. A. and Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review*, pages 267–271.

Harris, D. N. and Sass, T. R. (2014). Skills, productivity and the evaluation of teacher performance. *Economics of Education Review*, 40(0):183–204.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Hoxby, C. M. (2000). qthe effects of class size on student achievement: New evidence from population variation. *qQuarterly Journal of Economics,(2000a)*, 115(4):1239–85.

Hsieh, C.-T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from chile's voucher program. *Journal of public Economics*, 90(8):1477–1503.

Jacob, B. A. and Lefgren, L. (2008). Can principals identify effective teachers? evidence on subjective performance evaluation in education. *Journal of labor Economics*, 26(1):101–136.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? evidence from new york city. *Economics of Education Review*, 27(6):615–631.

Kane, T. J. and Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical Report 14607, National Bureau of Economic Research.

Kinsler, J. (2012). Assessing rothstein's critique of teacher value-added models. *Quantitative Economics*, 3(2):333–362.

Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *The Economic Journal*, 111(468):1–28.

Ladd, H. F. (2002). School vouchers: A critical view. *Journal of economic perspectives*, pages 3–24.

Ladd, H. F. and Walsh, R. P. (2002). Implementing value-added measures of school effectiveness: getting the incentives right. *Economics of Education review*, 21(1):1–17.

Leckie, G. and Goldstein, H. (2011). Understanding uncertainty in school league tables*. *Fiscal Studies*, 32(2):207–224.

Maddala, G. S., Trost, R. P., Li, H., and Joutz, F. (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business & Economic Statistics*, 15(1):90–100.

McCaffrey, D. F., Lockwood, J., Koretz, D., Louis, T. A., and Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of educational and behavioral statistics*, 29(1):67–101.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., and Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4(4):572–606.

McEwan, P. J. and Carnoy, M. (2000). The effectiveness and efficiency of private schools in chile's voucher system. *Educational evaluation and policy analysis*, 22(3):213–239.

McEwan, P. J., Urquiola, M., Vegas, E., Fernandes, R., and Gallego, F. A. (2008). School choice, stratification, and information on school performance: Lessons from chile [with comments]. *Economia*, pages 1–42.

Morris, C. N. (1983). Parametric empirical bayes inference: theory and applications. *Journal of the American Statistical Association*, 78(381):47–55.

Muralidharan, K. and Sheth, K. (2015). Bridging education gender gaps in developing countries: The role of female teachers. *Journal of Human Resources*.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, pages 1–32.

Patrinos, H. A. (2000). Market forces in education. *European Journal of Education*, pages 61–80.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69(2):167–190.

Rabe-Hesketh, S., Skrondal, A., and Zheng, X. . . (2007). *Multilevel structural equation modeling*, volume Handbook of Structural Equation Modeling. Elsevier.

Raudenbush, S. and Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of education*, pages 1–17.

Reardon, S. F. and Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, 4(4):492–519.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458.

Robinson, G. K. (1991). That blup is a good thing: the estimation of random effects. *Statistical science*, 6(1):pp. 15–32.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, pages 247–252.

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4):537–571.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1):pp. 175–214.

Rothstein, J. (2014). Revisiting the impacts of teachers. *Unpublished working paper. http://eml. berkeley. edu/˜ jrothst/workingpapers/rothstein_cfr. pdf.*

Sanders, W. L. and Horn, S. P. (1994). The tennessee value-added assessment system (tvaas): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in education*, 8(3):299–311.

Sanders, W. L. and Horn, S. P. (1998). Research findings from the tennessee value-added assessment system (tvaas) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3):247–256.

Sanders, W. L. and Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement.

Sapelli, C. and Vial, B. (2002). The performance of private and public schools in the chilean voucher system. *Cuadernos de economía*, 39(118):423–454.

Sass, T. R., Semykina, A., and Harris, D. N. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, 38(0):9–23.

Skrondal, A. and Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(3):659–687.

Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):pp. F3–F33.

Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *The American Economic Review*, 99(1):179–215.

Winters, M. A., Haight, R. C., Swaim, T. T., and Pickering, K. A. (2013). The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. *Economics of Education Review*, 34:69–75.

Chapter 8

Appendices

8.1 Appendix - Chapter 2

Appendix 2.1 Summary of Data Sets in the VAM Literature

Table 8.1: Datasets summary

Data Set	Authors	Period	Grades	Observations
North Carolina Education Research Data Centre	Clotfelter et al. 2006	2000-2001	5	Cross-section
	Clotfelter et al. 2007	1995-2004	3, 4, 5	Longitudinal data
	Rothstein 2009, 2010	2000-2001	3, 4, 5	Dropping classes with less than 12 students
				Dropping classes without parallel classes-grade
				Dropping students who change schools
	Jackson and Bruegman 2009	1995-2006	4, 5	Remove teachers who are co-teaching or have teaching aid
	Jackson 2012	1995-2006	3, 4, 5	Only classes with teachers who teach Language and Math to the same class
				Remove teachers who are co-teaching or have teaching aid
	Kinsler 2012		3, 4	Dropping classes with less than 15 students
				Dropping repeaters and school movers
Los Angeles Unified School District	Kane and Staiger 2008	2003-2005	2, 3, 4, 5	Dropping classes with less than 10 students
				Dropping classes with more than 36 students
	Buddin 2010, 2011	2002-2010	2, 3, 4, 5	Around 18,000 teachers
				520 schools
UTD Texas School Project	Rivkin et al. 2005	1993-1995	3, 4, 5, 6, 7	Around 200,000 students per cohort
				3,000 public elementary and middle schools
	Hanushek and Rivkin 2010a	1995-2001	4, 5, 6, 7, 8	Dropping students whitout valid test scores
				Dropping classes with less than 5 students
Chicago Public Schools	Aaronson et al. 2007	1996-1999	8, 9	Around 6,890 teachers
				1,132 9th Maths teachers
New York City Public Schools	Kane et al. 2008	1999-2005	4, 5, 6, 7, 8	Dropping classes with less than 7 students
				Dropping classes with more than 45 students
				Dropping classes where teacher reported working in more than one school
				Dropping students without test scores previous year
Single New Jersey County	Rockoff 2004	1989-2001/02	1, 2, 3, 4, 5, 6	District A/B
				Around 10,000 students
				300 teachers
Midsize School District Western US	Jacob and Lefgren 2008	1997-2005	2, 3, 4, 5, 6	Dropping Pre-school and 1st grade because non achievement exams
				Excluding non-core subjects teachers
				Aroun 200 teachers
Large Urban School District (US)	Chetty et al. 2014	1988-2009	4, 5, 6, 7, 8	Dropping classes with more than 25% of student with special education
				Dropping classes with less than 10 students
				Dropping classes with more than 50 students
				Dropping teachers linked to more than 200 students per grade
Dataset for the UK	Slater et al. 2012	1999-2002	7, 8, 9, 10, 11	Around 7,305 pupils
				740 teachers
				33 State secondary schools in Engalnd
Large and Diverse Anonymous State (US)	Dieterle et al. 2014	2001-2007	4, 5, 6	Dropping schools with less than 20 students per grade
				Dropping classes with less than 12 students
				Dropping districts with less than 1,000 students

Appendix 2.2 Transformation on the GAF

From equation (2.3), when $0 \leq \lambda \leq 1$ we have the GAF:

$$A_{i,g} = \sum_{r=1}^{g} \lambda^{g-r}[x'_{i,r}\beta_r + e'_{i,r}\gamma_r] + \alpha_i + \varepsilon_{i,g} \tag{8.1}$$

Specifying the GAF for a particular grade ($g = t$), we get the following expression:

$$A_{i,t} = \sum_{r=1}^{t} \lambda^{t-r}[x'_{i,r}\beta_r + e'_{i,r}\gamma_r] + \alpha_i + \varepsilon_{i,t} \tag{8.2}$$

Showing one period lag of equation (8.1), we have an expression for $g = t-1$:

$$A_{i,t-1} = \sum_{r=1}^{t-1} \lambda^{t-1-r}[x'_{i,r}\beta_r + e'_{i,r}\gamma_r] + \alpha_i + \varepsilon_{i,t-1} \tag{8.3}$$

If we multiply both sides of equation (8.3) by λ, we get:

$$\lambda A_{i,t-1} = \sum_{r=1}^{t-1} \lambda^{t-r}[x'_{i,r}\beta_r + e'_{i,r}\gamma_r] + \lambda\alpha_i + \lambda\varepsilon_{i,t-1} \tag{8.4}$$

Then, subtracting both sides of equation (8.2) by equation (8.4) we have that:

$$\begin{aligned} A_{i,t} - \lambda A_{i,t-1} = \sum_{r=1}^{t-1} \lambda^{t-r}[x'_{i,r}\beta_r + e'_{i,r}\gamma_r] + [x'_{i,t}\beta_t + e'_{i,t}\gamma_t] + \alpha_i + \varepsilon_{i,t} \\ - \sum_{r=1}^{t-1} \lambda^{t-r}[x'_{i,r}\beta_r + e'_{i,r}\gamma_r] - \lambda\alpha_i - \lambda\varepsilon_{i,t-1} \\ A_{i,t} - \lambda A_{i,t-1} = x'_{i,t}\beta_t + e'_{i,t}\gamma_t + (1-\lambda)\alpha_{i,t} + \varepsilon_{i,t} - \lambda\varepsilon_{i,t-1} \end{aligned}$$

$$A_{i,t} = \lambda A_{i,t-1} + x'_{i,t}\beta_t + e'_{i,t}\gamma_t + (1-\lambda)\alpha_i + \varepsilon_{i,g} - \lambda\varepsilon_{i,t-1} \tag{8.5}$$

and generalising for all g we get to the same equation (2.4) for **Model 1**.

$$A_{i,g} = \lambda A_{i,g-1} + x'_{i,g}\beta_g + e'_{i,g}\gamma_g + (1-\lambda)\alpha_{i,g} + \varepsilon_{i,g} - \lambda\varepsilon_{i,g-1} \tag{8.6}$$

8.2 Appendix - Chapter 3

Appendix 3.1 Cleaning process: Student panel 2003-2012

1. Identifying Good IDs and Bad IDs

Initial Panel	Freq.	Percent	Cum.
Good*	26,845,709	97%	98%
Bad**	**715,334**	**3%**	**100%**
Total	27,561,043	100%	

**Good: Observations belonging to IDs with one register per year through the period*

**Bad: Observations belonging to IDs who have at least one year with duplicated observations

2.1 From Bad Initial Panel

total_N	Freq.	Percent	Cum.	
2	681,309	95%	98%	
3 or More	**34,025**	**5%**	**100%**	***Dropped IDs***
Total	715,334	100%		

2.2 Total Observations of those Students who present Duplicated RBD in at least one yea **583,097**

We get to this number reshaping the 681,309 observations with at least one year with duplicated schools

3. Assignation Sequence for the observations in (2.2)

A. Observations without duplication problems:	
Observations with only one RBD (Years where the ID only has one RBD)	*484,885*
3.1 Total Observations with Assigned RBD	**98,212**
B. Solving backwards form the last year:	
RBD coincidence 2x2	
We choose randomly	*298*
Cases where we can find a pivot RBD	
If any of two RDB coincides with any of the previous year, we rule the macthed RDB	*957*
Cases where the match is not possible at all	
We choose randomly	*3,769*
C. Solving from the middle of the sample:	
We assign using backward criterion for school changes	*70,940*
D. Solving first year of the sample:	
We assign any following coincident RBD	*1,212*
E. Not possible relationships	
We choose randomly	*21,036*

Total observations with discresional assigned RBD	***98,212***
Percentage of assigned RBD on the final Panel	***0.4%***
Total observations which were randomly assigned	***25,103***
Percentage of randomly assigned RBD on the final Panel	***0.1%***
Total observations under RBD criteria selection	***583,097***
Percentage of total observations Student Panel	***2.1%***
Total Observations of Student Panel (2003-2012)	***27,428,806***

Appendix 3.2 Data sets availability - (2002-2013)

	Availability of Data Sets											
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Administrative data set												
Enrolment DB	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Performance DB	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Student Marks DB	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
School Directory DB	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teachers DB	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
School Staff DB	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
National Examination (Simce) data set												
Individual Scores	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Directories	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Municipalities	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Parental Questionnaire	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Teacher Questionnaire	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Other data bases												
Teaching Evaluation	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Vble. Pymnt. Indiv. Teacher Perfrm.	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Pedagogical Excellence Reward	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Educational Assistants	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No
Initial Evaluation	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes

Appendix 3.3 The administrative data set - Description tables

Variables description

Variable Name	Description
Year	Year
RBD	Key code for schools
NameRBD	Name of the school
CodRegRBD	Code of the region where the school is set. Regional Politic Division of Chile (1-13) and (1-15 From 2007)
CodMunRBD	Code of the municipality where the school is established
NameMunRBD	Name of the municipality where the school is established
CodDep	Type of school dependence (0 Municipal; 1 Private Voucher School; 2 Unsubsidised Private School)
RuralRBD	Geographic area where the school is set (0 Urban; 1 Rural)
CodEns	Primary, secondary or nursery schools
CodGrade	Grade from 1st to 8th *(Primary)* and from 9th to 12th *(Secondary)*
Letter	Identification class in the same grade *(e.g. from A to D, depends on the number of students)*
ClassComb	Type of class. Whether there are combined grades in the same class
Mrun	Unique identification number per student *(he/she keeps the same key number for the whole panel)*
GenMrun	Student gender (*0 Male; 1 Female*)
DateBirthMrun	Student date of birth
AgeMrun	Student age at June 30th of the corresponding year
ReptStd	Repeating student (*0 No; 1 Yes*)
DifGroup	Indicator whether the student is taught in a differential group (0 No; 1 Yes)
CodRegMrun	Code of the region where the student live
CodMunMrun	Code of the municipality where the student live
NameMunMrun	Name of the municipality where the student live
RcdEns	Classification recode of Primary, secondary and nursery school
SpcNeeds	Student with special transitories or permanent needs (*0 No; 1 Yes*)
Gpa	Grade point average *(Scale in Chile is from 1.0 to 7.0)*
Attendance	Percentage of attendance during the year
FinalStatus	Student final status: Promoted (P), Failed (F), Out (O)
FinalStatusTr	Student final status with transferred: Promoted (P), Failed (F), Out (O), Transferred (T)
Subject	Subject or teaching area (1 Language; 2 Maths; 3 Science; 4 Other)
Mark	Final mark in the specific subject (Language and Maths)
AddressRBD	School address
PhoneRBD	School phone number
Teacher_ID	Unique identify number per teacher
TeachingSubject	Teaching area (1 Language; 2 Maths; 3 Science; 4 Other)
Staff_ID	Unique identify number per staff (It is composed by principals, teachers and other pedagogical employees)
GenTeacher	Teacher gender (1 Male; 2 Female)
BirthTeacher	Teacher date of birth
TypeDegree	Type of education degree (11 Nursery; 12 Special needs; 13 Primary; 14 Scondary)
TypeDegreeEduc	Type of education degree (11 Nursery; 12 Special needs; 13 Primary; 14 Scondary)
PrimaryRole	Primary role of the staff (1 Classroom teacher; 3 Board member; 4 Principal or Head Master)
TypeContract	Type of contract depending on the type of school
HoursContract	Total working hours by contract
TenureEducSys	Years working in the educational system
TenureRBD	Years working in the school
TeachingSubject	Teaching subject (1 Language; 2 Maths; 3 Science; 4 Other)
TeachingHours	Total number of teaching hours

Variables availability

Variable Name	Enrolment DB	Performance DB	Student Marks DB	School Directory DB	Teachers DB	School Staff DB
Year	2004 - 2013	2002 - 2012	2003-09 / 2011-12	2004 - 2013	2002 - 2012	2003 - 2013
RBD	2004 - 2013	2002 - 2012	2003-09 / 2011-12	2004 - 2013	2002 - 2012	2003 - 2013
NameRBD	2004 - 2013	2002 - 2012	N/A	2004 - 2013	N/A	2003 - 2013
CodRegRBD	2004 - 2013	2002 - 2012	N/A	2004 - 2013	N/A	2003 - 2013
CodMunRBD	2004 - 2013	2002 - 2012	N/A	2004 - 2013	N/A	2003 - 2013
NameMunRBD	2004 - 2013	2002 - 2012	N/A	2004 - 2013	N/A	2003 - 2013
CodDep	2004 - 2013	2002 - 2012	N/A	2004 - 2013	N/A	2003 - 2013
RuralRBD	2004 - 2013	2002 - 2012	N/A	2004 - 2013	N/A	2003 - 2013
CodEns	2004 - 2013	2002 - 2012	2003-09 / 2011-12	2004 - 2013	2002 - 2012	N/A
CodGrade	2004 - 2013	2002 - 2012	2003-09 / 2011-12	N/A	2002 - 2012	N/A
Letter	2004 - 2013	2002 - 2012	2003-09 / 2011-12	N/A	N/A	N/A
ClassComb	2004 - 2013	N/A	N/A	N/A	N/A	N/A
Mrun	2004 - 2013	2002 - 2012	2003-09 / 2011-12	N/A	N/A	N/A
GenMrun	2004 - 2013	2002 - 2012	N/A	N/A	N/A	N/A
DateBirthMrun	2004 - 2013	2002 - 2012	N/A	N/A	N/A	N/A
AgeMrun	2004 - 2013	N/A	N/A	N/A	N/A	N/A
ReptStd	2004 - 2013	N/A	N/A	N/A	N/A	N/A
DifGroup	2004 - 2013	N/A	N/A	N/A	N/A	N/A
CodRegMrun	2004 - 2013	N/A	N/A	N/A	N/A	N/A
CodMunMrun	2004 - 2013	2002 - 2012	N/A	N/A	N/A	N/A
NameMunMrun	2004 - 2013	2002 - 2012	N/A	N/A	N/A	N/A
RcdEns	2004 - 2013	N/A	N/A	N/A	N/A	N/A
SpcNeeds	N/A	2002 - 2012	N/A	N/A	N/A	N/A
Gpa	N/A	2002 - 2012	N/A	N/A	N/A	N/A
Attendance	N/A	2002 - 2012	N/A	N/A	N/A	N/A
FinalStatus	N/A	2002 - 2012	N/A	N/A	N/A	N/A
FinalStatusTr	N/A	2002 - 2012	N/A	N/A	N/A	N/A
Subject	N/A	N/A	2003-09 / 2011-12	N/A	N/A	N/A
Mark	N/A	N/A	2003-09 / 2011-12	N/A	N/A	N/A
AddressRBD	N/A	N/A	N/A	2004 - 2013	N/A	N/A
PhoneRBD	N/A	N/A	N/A	2004 - 2013	N/A	N/A
Teacher_ID	N/A	N/A	N/A	N/A	2002 - 2012	2003 - 2013
TeachingSubject	N/A	N/A	N/A	N/A	2002 - 2012	2003 - 2013
Teacher_ID / Staff_ID	N/A	N/A	N/A	N/A	N/A	2004 - 2013
GenTeacher	N/A	N/A	N/A	N/A	N/A	2003 - 2013
BirthTeacher	N/A	N/A	N/A	N/A	N/A	2003 - 2013
TypeDegree	N/A	N/A	N/A	N/A	N/A	2003 - 2013
TypeDegreeEduc	N/A	N/A	N/A	N/A	N/A	2003 - 2013
PrimaryRole	N/A	N/A	N/A	N/A	N/A	2003 - 2013
TypeContract	N/A	N/A	N/A	N/A	N/A	2003 - 2013
HoursContract	N/A	N/A	N/A	N/A	N/A	2003 - 2013
TenureEducSys	N/A	N/A	N/A	N/A	N/A	2003 - 2013
TenureRBD	N/A	N/A	N/A	N/A	N/A	2011 - 2013
TeachingArea	N/A	N/A	N/A	N/A	N/A	2003 - 2013
TeachingHours	N/A	N/A	N/A	N/A	N/A	2003 - 2013

Notes: (i) The definition of each variable is described in the previous table description. (ii) **N/A:** Not Available

Appendix 3.4 Teacher Questionnaire - Processed Variables Set

Variables Description

Language & Maths Teachers Questonnaire - From 2003 to 2009 (*Encuesta de Profesores*)	
RBD	**School ID: key code for schools**
CodGrado	**Grade from 1st to 8th *(Primary)* and from 9th to 12th *(Secondary)***
Letter	**Identification class in the same grade *(e.g. from A to D, depends on the number of students)***
TchrAge	Teacher's age (0 less than 31; 1 from 31 to 40; 2 from 41 to 50; 3 more than 50)
TchrGender	Teacher's gender (0 masculine, 1 female)
TchrExp	Teacher's experience (0 less than 5; 1 from 5 to 13; 2 from 14 to 20, 3 more than 20)
ThcrDegree	Teacher's degree (0 does not hold; 1 holds)
AddMthSpec	Additional maths specialization (0 no; 1 yes)
Wkhrs_sch	Working hours per week at the school (0 less than 26; 1 from 26 to 30; 2 from 30 to 40; 3 more than 40)
Teach_hrs	Teaching hours per week (0 less than 21, 1 from 21 to 30; 2 from 30 to 40)
Prep_hrs	Classes preparation hours per week (0 less than 3, 1 from 3 to 4; 2 from 5 to 10; 3 more than 10)
Plannif_time	Does the school assign ofical hours for plannification and class preparation? (0 no; 1 yes)
Teamwork	Do you work with other colleges for class preparation? (0 no; 1 yes)
Otherschool	Do you work in other school? (0 no; 1 yes)
Expstudatt	Expectation about students' attainment (0 Inc Prim; 1 Comp Prim; 2 Inc Sec; 3 Comp Sec; 4 Comp College, 5 Comp Uni)
Frequence of specific activities for teachers	
grwkstr	Use strategies for teamwork (0 never; 1 occasionally; 2 often; 3 always)
idvwkstr	Use strategies for individual work (0 never; 1 occasionally; 2 often; 3 always)
prevblcnt	Presents verbally learning contents (0 never; 1 occasionally; 2 often; 3 always)
orgclssqtan	Organise classes based on question and answers (0 never; 1 occasionally; 2 often; 3 always)
wrkwtprbsets	Work with problem sets (0 never; 1 occasionally; 2 often; 3 always)
wrkwtshortqts	Work with short questionaries (0 never; 1 occasionally; 2 often; 3 always)
askstuorptts	Ask stuidents for oral presentations (0 never; 1 occasionally; 2 often; 3 always)
askstursdhwtrep	Ask students for brief research and writing reports (0 never; 1 occasionally; 2 often; 3 always)
Frequence of specific activities for teachers (Language only)	
setdramplays	Set drama plays (0 never; 1 occasionally; 2 often; 3 always)
setdeborforu	Set debates or forums (0 never; 1 occasionally; 2 often; 3 always)
orgschexc	Organise school excursions (0 never; 1 occasionally; 2 often; 3 always)
Frequence of specific activities for teachers (Mathematics only)	
showmathsdem	Show mathematics demonstrations (0 never; 1 occasionally; 2 often; 3 always)
Frequency using different educational resources	
usetextbook	Usage of textbook (0 never; 1 occasionally; 2 often; 3 always)
usevideos	Usage of videos (0 never; 1 occasionally; 2 often; 3 always)
useeducsoft	Usage of educational softwares (0 never; 1 occasionally; 2 often; 3 always)
useinternet	Usage of internet (0 never; 1 occasionally; 2 often; 3 always)
Frequency using different educational resources (Language only)	
usedictionaries	Usage of dictionaries (0 never; 1 occasionally; 2 often; 3 always)
Frequency using different educational resources (Mathematics only)	
usecalcula	Usage of calculator (0 never; 1 occasionally; 2 often; 3 always)
measureinstr	Usage of measuring instruments (0 never; 1 occasionally; 2 often; 3 always)
Use of assessment tools	
obsrecords	Use of observation records (0 no; 1 yes)
oralexams	Use of oral exams (0 no; 1 yes)
tftestaorgfq	Use of true-false or fill-in-the-blank questions (0 no; 1 yes)
multchoictest	Use of multiple choice test (0 no; 1 yes)
writtentest	Use of written test (0 no; 1 yes)
Evaluation activities	
diagtestwomark	Diagnostic test without mark for every unit (0 no; 1 yes)
diagtestwithmark	Diagnostic test with mark for every unit (0 no; 1 yes)
midtestwomark_eu	Midterm test without mark for every unit (0 no; 1 yes)
midtestwithmark_eu	Midterm test with mark for every unit (0 no; 1 yes)
fintermtest	Final term test (0 no; 1 yes)
yearlyfinex	Yearly final exam (0 no; 1 yes)
corrpsandassi	Correct problem sets and assignments (0 no; 1 yes)
diswithstevres	Discuss with students evaluation results (0 no; 1 yes)
restosuplwperfst	Use results to support low performance students (0 no; 1 yes)
restomodplan	Use results to modify plannification classes (0 no; 1 yes)
Use the orientation provided	
oribrousimce	Use orientation brochure for preparing SIMCE Exam (0 no; 1 yes)

Note: We use the RBD, CodGrade and Letter variables (school ID, grade and class letter) to match with the rest of the student panel data set (SPD).

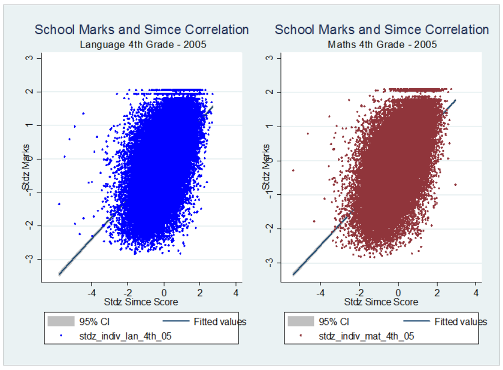
Variables Availability

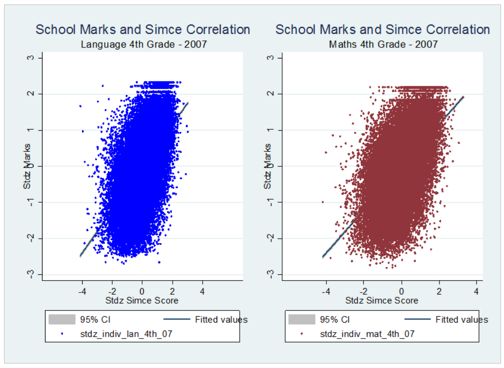
Availability of Variables in the Data Set by Year/Grade											
Year	**2003**	**2004**	**2005**	**2006**		**2007**		**2008**		**2009**	
Grade	**10th**	**8th**	**4th**	**4th**	**10th**	**4th**	**8th**	**4th**	**10th**	**4th**	**8th**
RBD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CodGrado	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Letter	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TchrAge	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TchrGender	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TchrExp	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
ThcrDegree	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
AddMthSpec	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Wkhrs_sch	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Teach_hrs	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Prep_hrs	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Plannif_time	Yes	Yes	Yes	No	No	No	Yes	No	No	No	No
Teamwork	Yes	No	No	No	No	No	No	No	No	No	No
Otherschool	Yes	Yes	Yes	No	No	No	Yes	No	Yes	No	No
Expstudatt	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Frequence of specific activities for teachers											
grwkstr	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
idvwkstr	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
prevblcnt	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
orgclssqtan	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
wrkwtprbsets	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
wrkwtshortqts	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
askstuorptts	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
askstursh wtrep	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
Frequence of specific activities for teachers (Language only)											
setdramplays	Yes	Yes	No	No	No	No	No	No	No	No	No
setdeborforu	Yes	Yes	No	No	No	No	No	No	No	No	No
orgschexc	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	No
Frequence of specific activities for teachers (Mathematics only)											
showmathsdem	Yes	No	No	No	No	No	No	No	No	No	No
Frequence using different educational resources											
usetextbook	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No
usevideos	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No
useeducsoft	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No
useinternet	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No
Frequence using different educational resources (Language only)											
usedictionaries	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No
Frequence using different educational resources (Mathematics only)											
usecalcula	Yes	Yes	Yes	No	No	No	Yes	No	Yes	No	No
measureinstr	Yes	No	No	No	No	No	No	No	No	No	No
Use of assessment tools											
obsrecords	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
oralexams	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
tftestaorgfq	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
multchoictest	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
writtentest	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
Evaluation activities											
diagtestwomark	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
diagtestwithmark	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
midtestwomark_eu	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
midtestwithmark_eu	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
fintermtest	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
yearlyfinex	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
corrpsandassi	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
diswithstevres	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
restosuplwperfst	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
restomodplan	Yes	Yes	No	No	No	No	Yes	No	Yes	No	No
Use the orientation provided											
oribrousimce	Yes	No	Yes	Yes	Yes	No	No	No	No	No	No

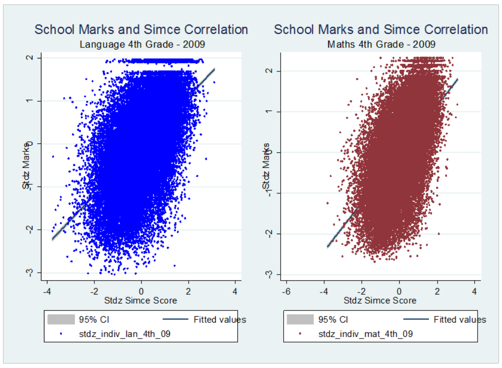
Note: The definition of each variable is described in the previous table description.

Appendix 3.5 Correlation graphs

Scatter plots







8.3 Appendix - Chapter 4

Appendix 4.1 Mini panels description and sample selection

Table 8.2: Summary Statistics - Cohort 2

	Cohort 2				Group 1				Group 2			
Variable	**Mean**	**Std.Dev.**	**Min**	**Max**	**Mean**	**Std.Dev.**	**Min**	**Max**	**Mean**	**Std.Dev.**	**Min**	**Max**
Pupil Level												
GPA	5.66	0.73	1	7	5.84	0.57	2.3	7	5.84	0.55	2.8	7
Average school Language marks	5.38	0.75	1	7	5.50	0.70	1.8	7	5.47	0.69	3.0	7
Average school Maths marks	5.29	0.81	1	7	5.40	0.77	2.7	7	5.39	0.77	2.7	7
Gender (Female=1)	0.49	0.50	0	1	0.50	0.50	0	1	0.50	0.50	0	1
Age	9.79	0.75	7	15	9.70	0.60	7	15	9.71	0.58	7	15
Attendance	92.56	9.56	0	100	93.75	6.56	0	100	93.86	5.63	28.3	100
Special Needs	0.02	0.11	0	1	0.01	0.10	0	1	0.01	0.08	0	1
Class Level												
Available Language teacher	0.98	0.14	0	1	0.98	0.13	0	1	0.97	0.17	0	1
Available Maths teacher	0.98	0.13	0	1	0.99	0.11	0	1	0.97	0.16	0	1
No. Schools per Language teacher	1.03	0.18	1	3	1.04	0.20	1	3	1.04	0.20	1	2
No. Schools per Maths teacher	1.03	0.18	1	3	1.04	0.20	1	3	1.03	0.17	1	2
No. Grades per Language teacher	2.48	1.71	1	8	1.63	0.95	1	7	1.55	0.81	1	6
No. Grades per Maths teacher	2.55	1.75	1	8	1.70	1.05	1	8	1.55	0.83	1	6
No. Classes per Language teacher	2.48	1.71	1	8	1.63	0.95	1	7	1.55	0.81	1	6
No. Classes per Maths teacher	2.91	1.94	1	12	2.31	1.81	1	10	2.48	1.85	1	9
Subject Specialist teacher	0.33	0.47	0	1	0.41	0.49	0	1	0.44	0.50	0	1
School Level												
Municipal schools	0.58	0.49	0	1	0.46	0.50	0	1	0.49	0.50	0	1
Private Voucher schools	0.37	0.48	0	1	0.47	0.50	0	1	0.38	0.49	0	1
Unsubsidised Private schools	0.05	0.22	0	1	0.07	0.25	0	1	0.13	0.33	0	1
Rural Area	0.47	0.50	0	1	0.04	0.19	0	1	0.01	0.09	0	1
School Socioeconomic Level	1.19	1.17	0	4	2.00	0.95	0	4	2.20	0.96	0.3	4
Number of students per grade	29.98	34.78	1	469	66.47	12.53	35.7	93	104.79	16.81	60.6667	138
Number of students per class	18.50	13.87	1	47	26.02	6.06	15	45	28.14	5.38	17.8	43
Number of students	257,344				60,224				26,517			
Number of classes	33,123				7,008				2,862			
Number of schools	8,500				1,168				318			

Notes: (i) Group 1: Schools with 2 classes per grade from 3rd to 5th grade; **Group 2:** Schools with 3 classes per grade from 3rd to 5th grade. **(ii)** The number of students corresponds to the total pupil observed in Groups 1 and 2. **(iii)** The number of classes corresponds to all the classrooms observed over the three year periods in all schools. **(iv)** The number of schools refers to the total schools observed over the panel, then if a pupil from the reference cohort (4th grade 2005) is in a different school in 2004 or 2006, that school it would be counted here. **(v)** School Socioeconomic Level variable: 0 Low; 1 Mid Low; 2 Middle; 3 Middle High; 4 High.

Table 8.3: Summary Statistics - Cohort 3

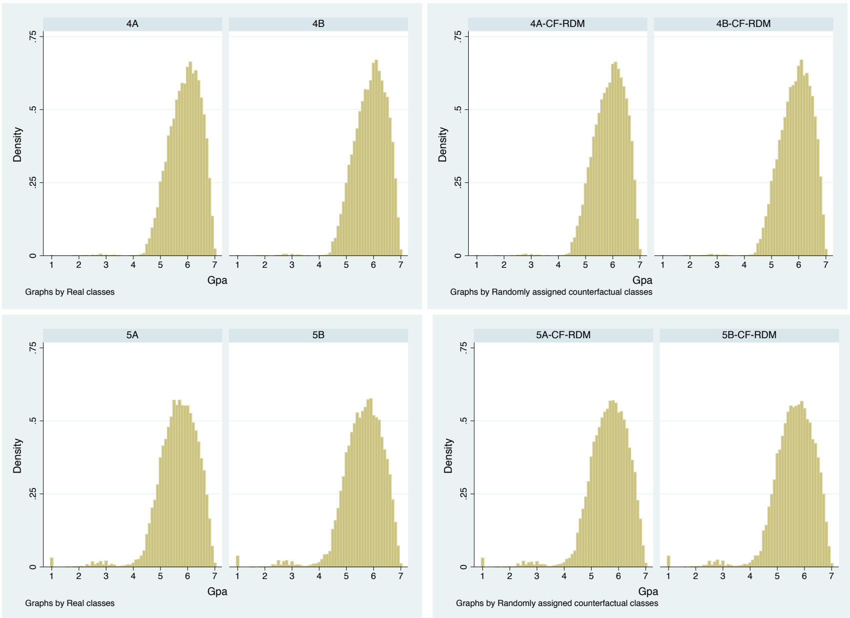
Variable	Cohort 3 Mean	Std.Dev.	Min	Max	Group 1 Mean	Std.Dev.	Min	Max	Group 2 Mean	Std.Dev.	Min	Max
Pupil Level												
GPA	5.63	0.74	1	7	5.81	0.56	1.0	7	5.81	0.55	2.7	7
Average school Language marks	5.36	0.73	1	7	5.46	0.69	1.0	7	5.45	0.67	3.3	7
Average school Maths marks	5.27	0.79	1	7	5.36	0.77	1.0	7	5.37	0.76	3.0	7
Gender (Female=1)	0.49	0.50	0	1	0.51	0.50	0	1	0.48	0.50	0	1
Age	9.82	0.76	7	15	9.72	0.60	8	14.5	9.73	0.58	8.7	14.5
Attendance	91.14	11.03	0	100	93.05	6.13	0	100	93.18	5.92	15.3	100
Special Needs	0.01	0.09	0	1	0.01	0.07	0	1	0.01	0.06	0	1
Class Level												
Available Language teacher	0.99	0.12	0	1	0.99	0.08	0	1	0.99	0.11	0	1
Available Maths teacher	0.99	0.10	0	1	1.00	0.06	0	1	0.99	0.10	0	1
No. Schools per Language teacher	1.02	0.15	1	3	1.03	0.16	1	2	1.03	0.16	1	2
No. Schools per Maths teacher	1.03	0.16	1	2	1.03	0.17	1	2	1.03	0.16	1	2
No. Grades per Language teacher	2.49	1.68	1	8	1.71	1.00	1	8	1.59	0.81	1	5
No. Grades per Maths teacher	2.58	1.73	1	8	1.78	1.07	1	6	1.59	0.86	1	6
No. Classes per Language teacher	2.49	1.68	1	8	1.71	1.00	1	8	1.59	0.81	1	5
No. Classes per Maths teacher	2.98	1.93	1	13	2.51	1.85	1	10	2.71	1.94	1	10
Subject Specialist teacher	0.38	0.49	0	1	0.49	0.50	0	1	0.52	0.50	0	1
School Level												
Municipal schools	0.57	0.50	0	1	0.42	0.49	0	1	0.45	0.50	0	1
Private Voucher schools	0.38	0.49	0	1	0.51	0.50	0	1	0.42	0.49	0	1
Unsubsidised Private schools	0.05	0.22	0	1	0.07	0.26	0	1	0.13	0.34	0	1
Rural Area	0.46	0.50	0	1	0.04	0.19	0	1	0.01	0.10	0	1
School Socioeconomic Level	1.23	1.12	0	4	1.99	0.98	0	4	2.18	1.02	0.0	4
Number of students per grade	29.50	33.34	1	379	66.18	12.61	34.3	92	103.56	17.05	60.7	136
Number of students per class	18.63	13.75	1	47	26.22	6.20	15	44	27.87	5.56	17.4	43
Number of students	250,275				61,086				24,780			
Number of classes	32,574				70,049				2,699			
Number of schools	8,428				1,175				300			

Notes: (i) Group 1: Schools with 2 classes per grade from 3rd to 5th grade; **Group 2:** Schools with 3 classes per grade from 3rd to 5th grade. **(ii)** The number of students corresponds to the total pupil observed in Groups 1 and 2. **(iii)** The number of classes corresponds to all the classrooms observed over the three year periods in all schools. **(iv)** The number of schools refers to the total schools observed over the panel, then if a pupil from the reference cohort (4th grade 2005) is in a different school in 2004 or 2006, that school it would be counted here. **(v)** School Socioeconomic Level variable: 0 Low; 1 Mid Low; 2 Middle; 3 Middle High; 4 High.

Appendix 4.2 Graphic analysis

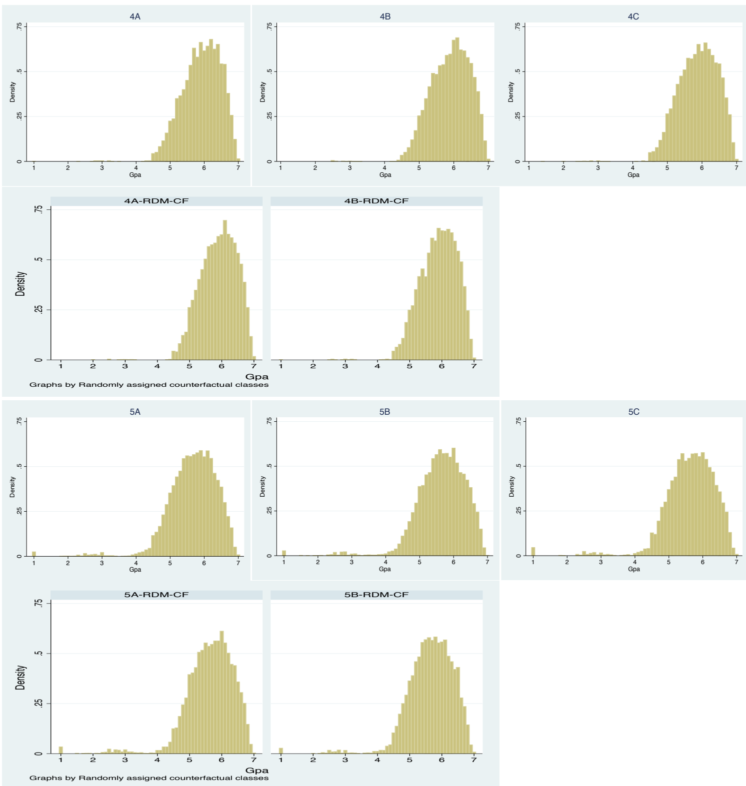
Cohort 2

Figure 8.1: Real classes vs random counterfactuals based on GPA
Group 1 (Cohort 2)



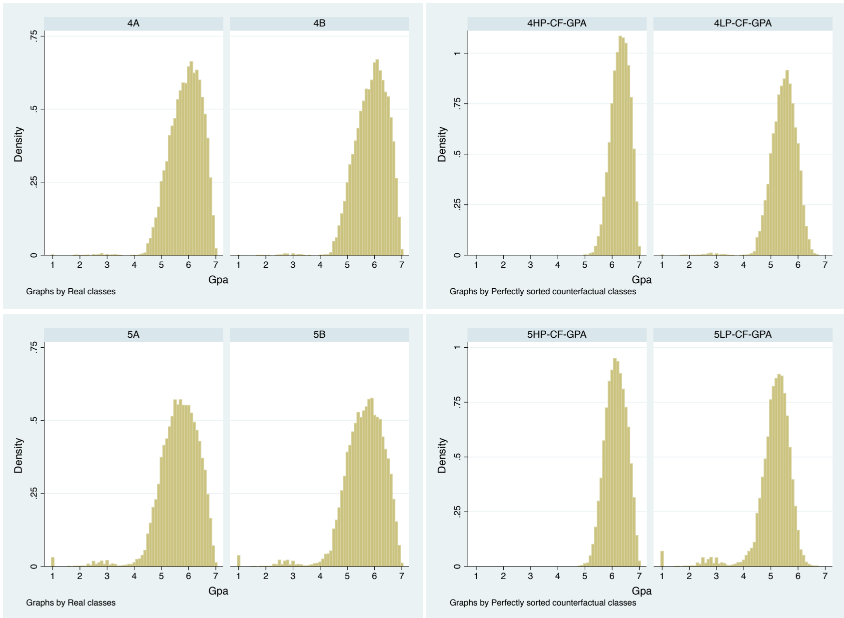
Note: RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes.

Figure 8.2: Real classes vs random counterfactuals based on GPA
Group 2 (Cohort 2)



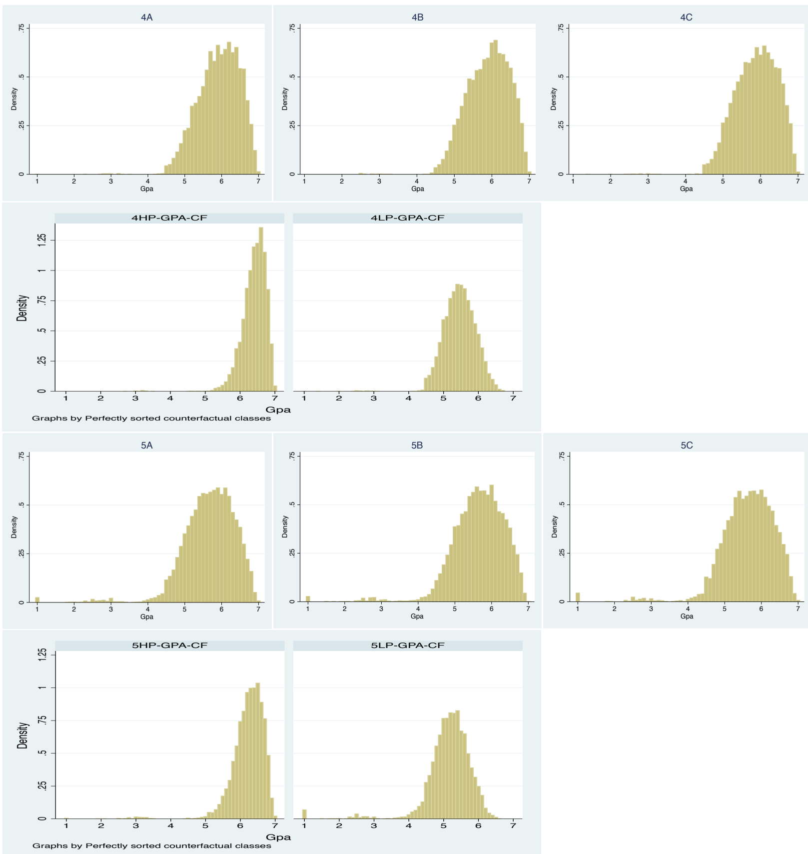
Note: RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes.

Figure 8.3: Real classes vs perfectly sorted counterfactuals based on GPA Group 1 (Cohort 2)



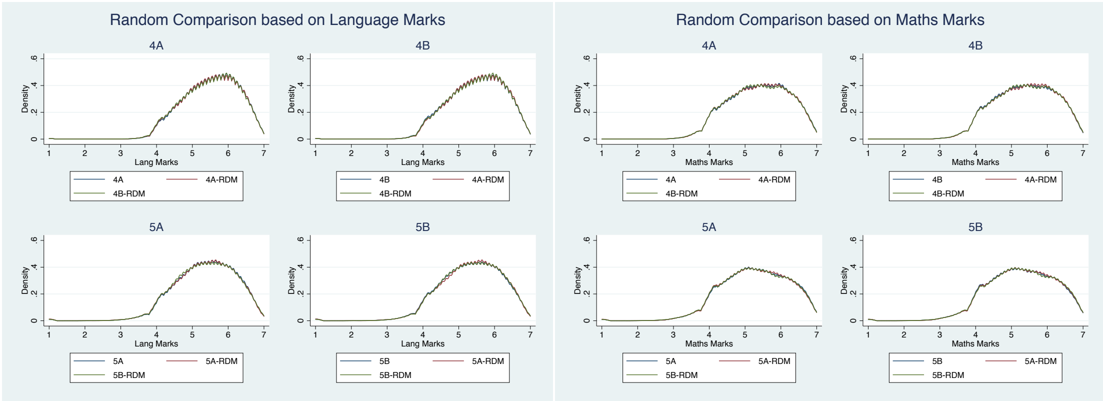
Notes: There are two perfectly sorted counterfactual classes per grade; (i) **HP-GPA-CF:** High Performance classes based on GPA. **(ii) LP-GPA-CF:** Low Performance classes based on GPA.

Figure 8.4: Real classes vs perfectly sorted counterfactuals based on GPA Group 2 (Cohort 2)



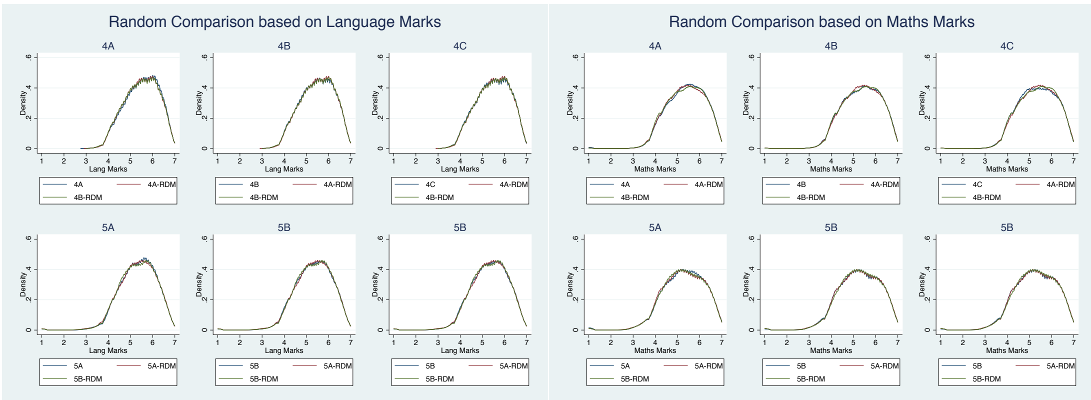
Notes: There are two perfectly sorted counterfactual classes per grade; (i) **HP-GPA-CF:** High Performance classes based on GPA. **(ii) LP-GPA-CF:** Low Performance classes based on GPA.

Figure 8.5: Real classes vs random counterfactuals based previous school marks Group 1 (Cohort 2)



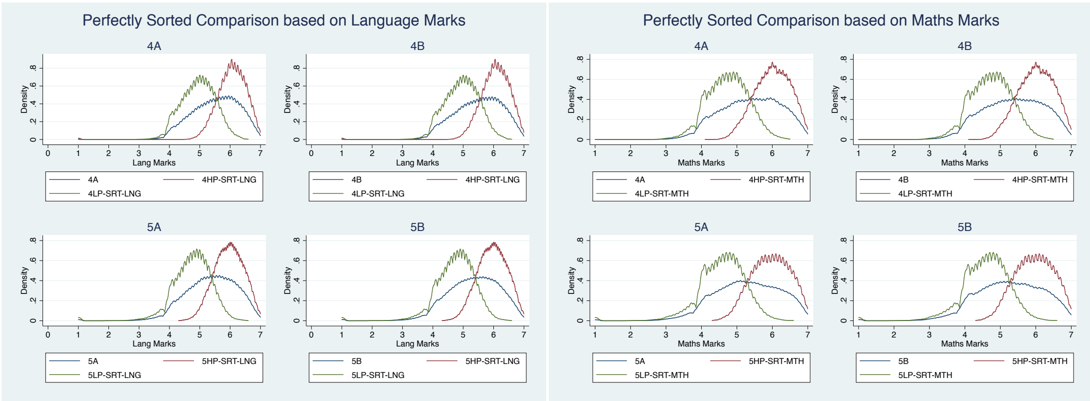
Notes: (i) RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes. **(ii)** In each plot we compare real classes **(4A, 4B, 5A, 5B)** with their respective random counterfactuals **(A, B RDM-CF).**

Figure 8.6: Real classes vs random counterfactuals based previous school marks Group 2 (Cohort 2)



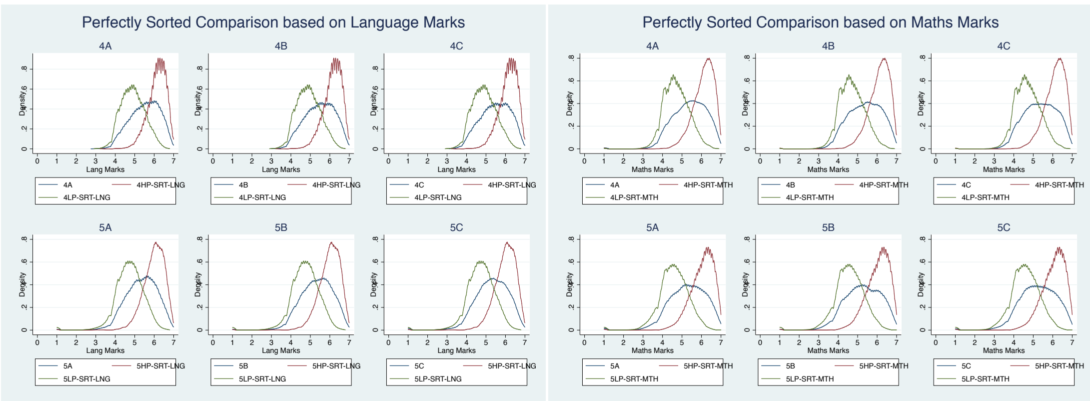
Notes: (i) RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes. **(ii)** In each plot we compare real classes **(4A, 4B, 4C, 5A, 5B, 5C)** with their respective random counterfactuals **(A, B RDM-CF).**

Figure 8.7: Real classes vs perfectly sorted counterfactuals based previous school marks
Group 1 (Cohort 2)



Notes: There are two perfectly sorted counterfactual classes per and grade. **(i)** Based on Language Marks; **HP-SRT-LNG:** High Performance classes; **LP-SRT-LNG:** Low Performance classes. **(ii)** Based on Maths Marks; **HP-SRT-MTH:** High Performance classes; **LP-SRT-MTH:** Low Performance classes. **(iii)** In each plot we compare real classes **(4A, 4B, 5A, 5B)** with their respective random counterfactuals **(HP-SRT-LNG, HP-SRT-MTH; LP-SRT-LNG, LP-SRT-MTH).**

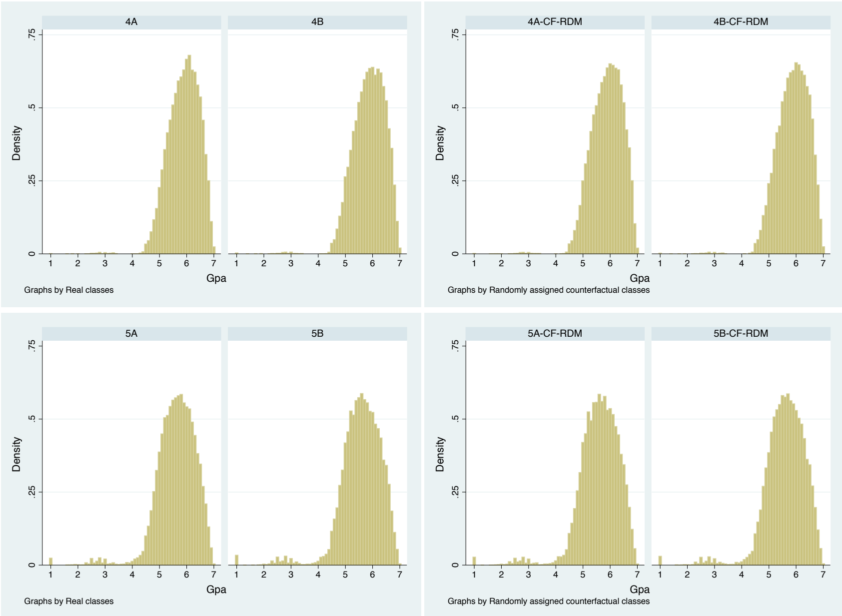
Figure 8.8: Real classes vs perfectly sorted counterfactuals based previous school marks
Group 2 (Cohort 2)



Notes: There are two perfectly sorted counterfactual classes per and grade. **(i)** Based on Language Marks; **HP-SRT-LNG:** High Performance classes; **LP-SRT-LNG:** Low Performance classes. **(ii)** Based on Maths Marks; **HP-SRT-MTH:** High Performance classes; **LP-SRT-MTH:** Low Performance classes. **(iii)** In each plot we compare real classes **(4A, 4B, 4C, 5A, 5B, 5C)** with their respective random counterfactuals **(HP-SRT-LNG, HP-SRT-MTH; LP-SRT-LNG, LP-SRT-MTH).**

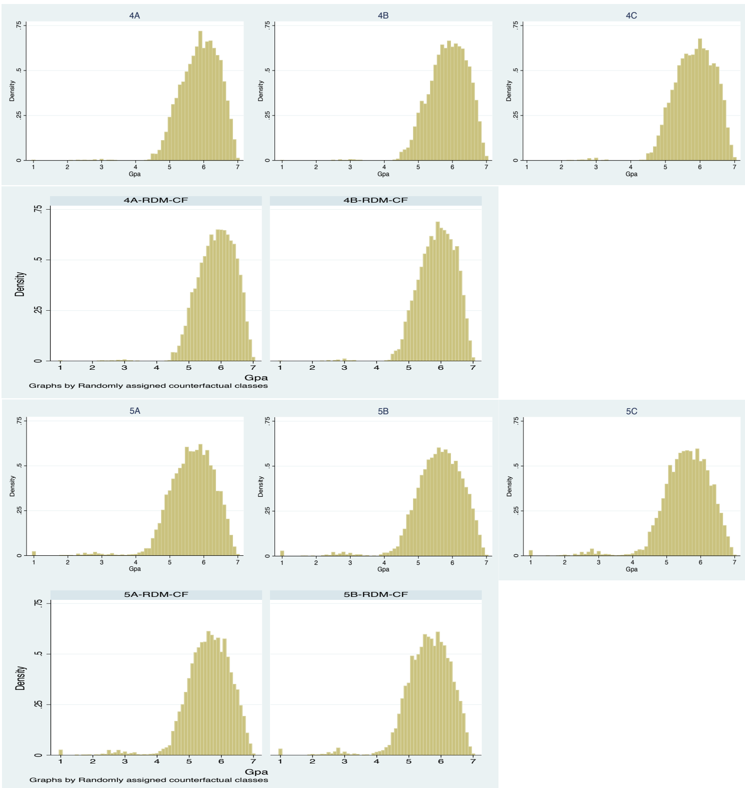
Cohort 3

Figure 8.9: Real classes vs random counterfactuals based on GPA
Group 1 (Cohort 3)



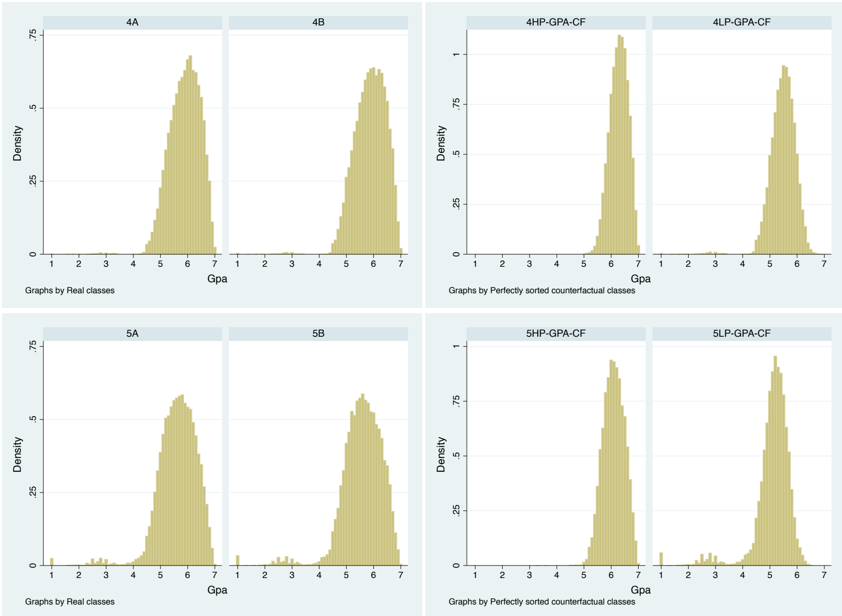
Note: RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes.

Figure 8.10: Real classes vs random counterfactuals based on GPA
Group 2 (Cohort 3)



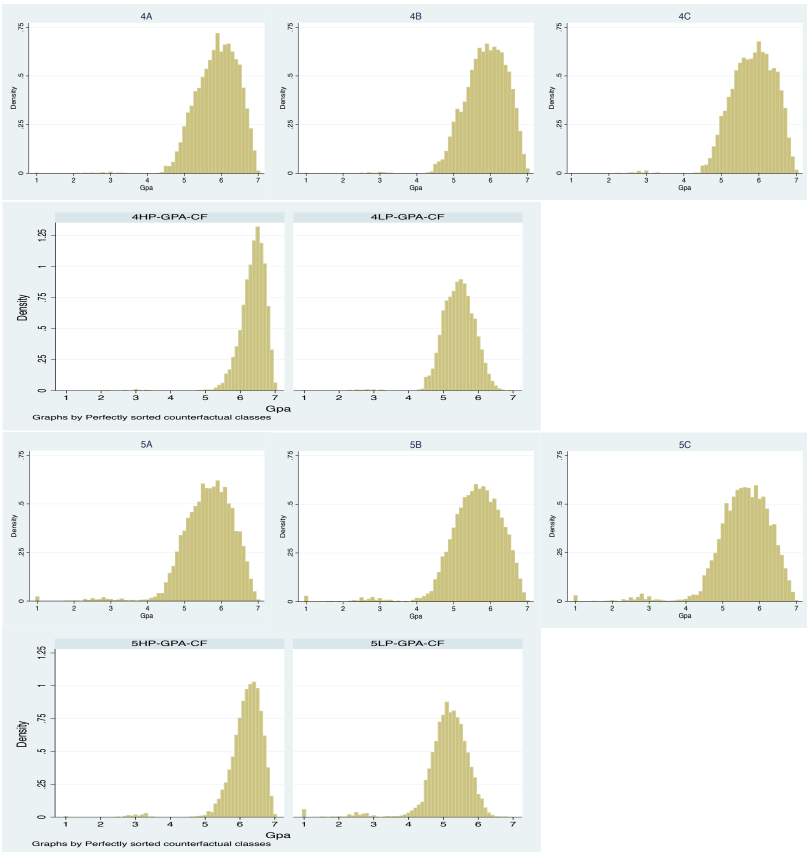
Note: RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes.

Figure 8.11: Real classes vs perfectly sorted counterfactuals based on GPA Group 1 (Cohort 3)



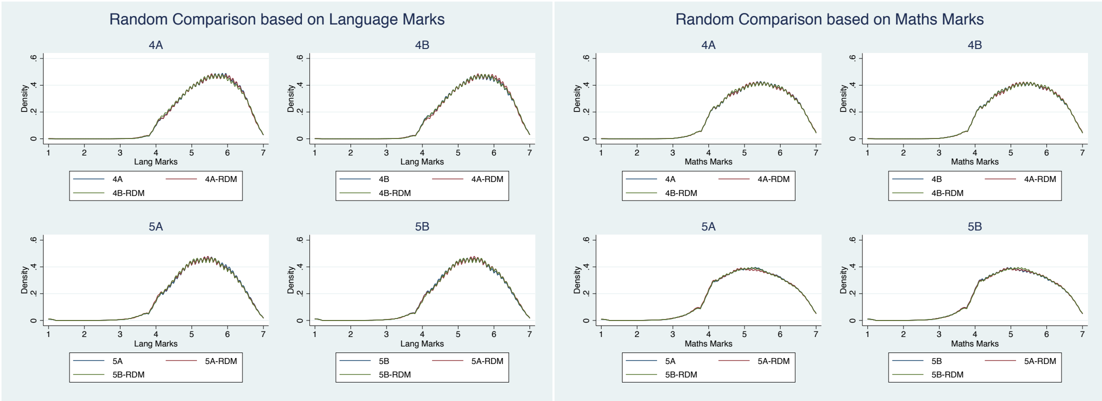
Notes: There are two perfectly sorted counterfactual classes per grade; (i) **HP-GPA-CF:** High Performance classes based on GPA. **(ii) LP-GPA-CF:** Low Performance classes based on GPA

Figure 8.12: Real classes vs perfectly sorted counterfactuals based on GPA Group 2 (Cohort 3)



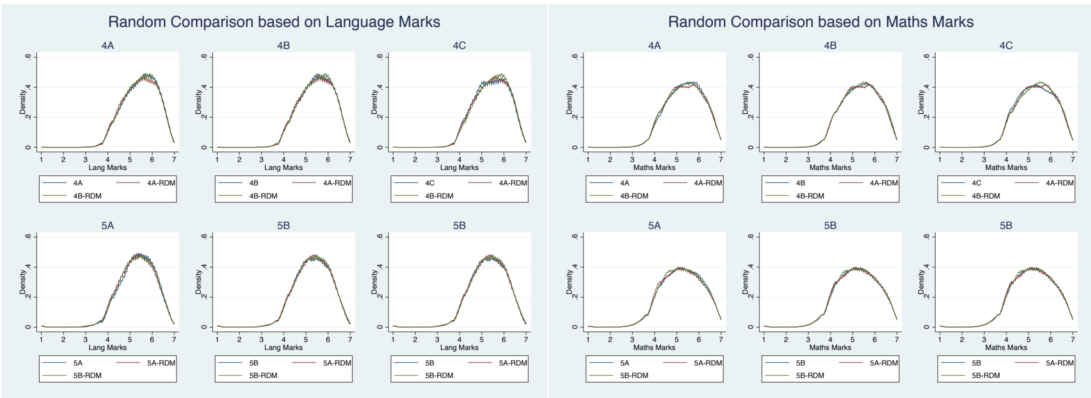
Notes: There are two perfectly sorted counterfactual classes per grade; (i) **HP-GPA-CF:** High Performance classes based on GPA. **(ii) LP-GPA-CF:** Low Performance classes based on GPA

Figure 8.13: Real classes vs random counterfactuals based previous school marks Group 1 (Cohort 3)



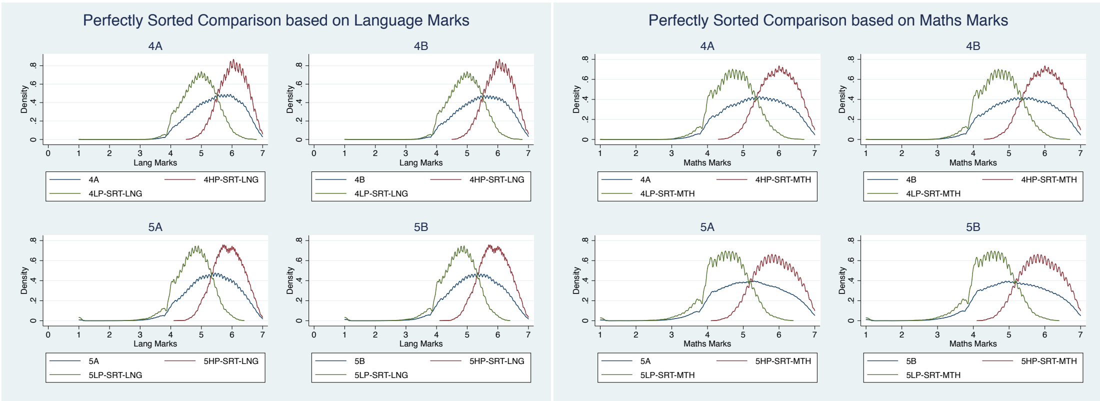
Notes: (i) RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes. **(ii)** In each plot we compare real classes **(4A, 4B, 5A, 5B)** with their respective random counterfactuals **(A, B RDM-CF).**

Figure 8.14: Real classes vs random counterfactuals based previous school marks Group 2 (Cohort 3)



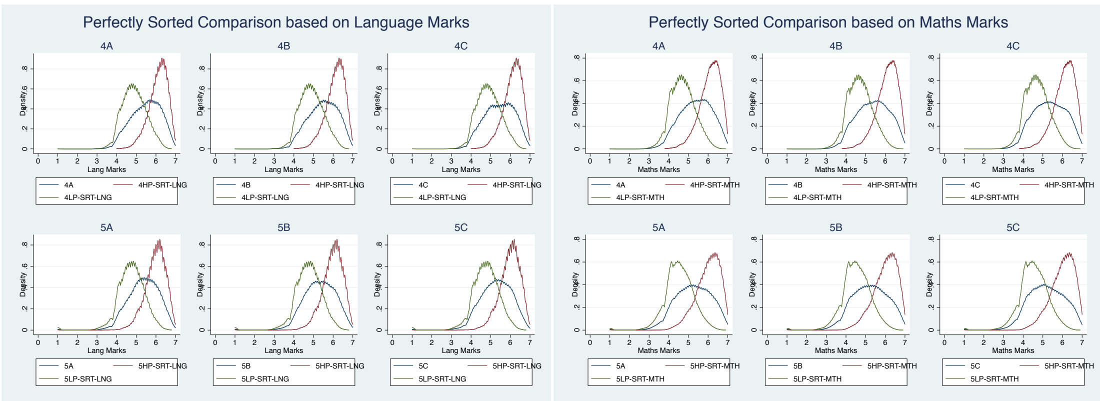
Notes: (i) RDM-CF: Randomly sorted counterfactuals. There are two counterfactual classes **(A, B)** per grade with pupils uniformly distributed into classes. **(ii)** In each plot we compare real classes **(4A, 4B, 4C, 5A, 5B, 5C)** with their respective random counterfactuals **(A, B RDM-CF).**

Figure 8.15: Real classes vs perfectly sorted counterfactuals based previous school marks
Group 1 (Cohort 3)



Notes: There are two perfectly sorted counterfactual classes per and grade. **(i)** Based on Language Marks; **HP-SRT-LNG:** High Performance classes; **LP-SRT-LNG:** Low Performance classes. **(ii)** Based on Maths Marks; **HP-SRT-MTH:** High Performance classes; **LP-SRT-MTH:** Low Performance classes. **(iii)** In each plot we compare real classes **(4A, 4B, 5A, 5B)** with their respective random counterfactuals **(HP-SRT-LNG, HP-SRT-MTH; LP-SRT-LNG, LP-SRT-MTH).**

Figure 8.16: Real classes vs perfectly sorted counterfactuals based previous school marks
Group 2 (Cohort 3)



Notes: There are two perfectly sorted counterfactual classes per and grade. **(i)** Based on Language Marks; **HP-SRT-LNG:** High Performance classes; **LP-SRT-LNG:** Low Performance classes. **(ii)** Based on Maths Marks; **HP-SRT-MTH:** High Performance classes; **LP-SRT-MTH:** Low Performance classes. **(iii)** In each plot we compare real classes **(4A, 4B, 4C, 5A, 5B, 5C)** with their respective random counterfactuals **(HP-SRT-LNG, HP-SRT-MTH; LP-SRT-LNG, LP-SRT-MTH).**

Appendix 4.3 Measuring Non-random assignment

Table Results - Cohort 2

Table 8.4: Estimated levels of Non-Random assignment
Group 1 & 2 (Cohort 2)

Non-RA Indexes				Sorting Evidence (SoE)					Total No. of schools
				None	Low	Medium	Med-High	High	
Random (RDM)	T-test	GPA	*Non-RA Index (1)*	1,169	266	41	9	1	1,486
		Language	*Non-RA Index (2)*	1,106	303	65	10	2	1,486
		Maths	*Non-RA Index (3)*	1,130	298	52	6	0	1,486
	KS-test	GPA	*Non-RA Index (4)*	1,363	108	12	3	0	1,486
		Language	*Non-RA Index (5)*	1,307	163	13	3	0	1,486
		Maths	*Non-RA Index (6)*	1,333	146	6	1	0	1,486
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*	711	668	103	4	0	1,486
		Language	*Non-RA Index (8)*	817	578	90	1	0	1,486
		Maths	*Non-RA Index (9)*	919	501	64	1	1	1,486
	KS-test	GPA	*Non-RA Index (10)*	731	536	214	5	0	1,486
		Language	*Non-RA Index (11)*	621	590	269	6	0	1,486
		Maths	*Non-RA Index (12)*	706	558	214	8	0	1,486
			Non-RA Average Index	**993**	**393**	**95**	**5**	**0**	

Notes: **(i)** In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools (ii) There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test. **(v)** The level of Non-Random assignment depends on the **Sorting Evidence (SoE)** in every school: None if **SoE**=0; Low if **SoE**=1,2 (1,2,3); Medium if **SoE**=3,4 (5,4,6); Med-High if **SoE**=5,6 (7,8,9); and High if **SoE**=7,8 (10,11,12); in brackets conditions for **SoE** categories in Group 2.

Table 8.5: Estimated levels of Non-Random assignment in percentage
Group 1 & 2 (Cohort 2)

Non-RA Indexes				Sorting Evidence (SoE)					Total schools (%)
				None	Low	Medium	Med-High	High	
Random (RDM)	T-test	GPA	*Non-RA Index (1)*	79%	18%	3%	1%	0%	100%
		Language	*Non-RA Index (2)*	74%	20%	4%	1%	0%	100%
		Maths	*Non-RA Index (3)*	76%	20%	3%	0%	0%	100%
	KS-test	GPA	*Non-RA Index (4)*	92%	7%	1%	0%	0%	100%
		Language	*Non-RA Index (5)*	88%	11%	1%	0%	0%	100%
		Maths	*Non-RA Index (6)*	90%	10%	0%	0%	0%	100%
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*	48%	45%	7%	0%	0%	100%
		Language	*Non-RA Index (8)*	55%	39%	6%	0%	0%	100%
		Maths	*Non-RA Index (9)*	62%	34%	4%	0%	0%	100%
	KS-test	GPA	*Non-RA Index (10)*	49%	36%	14%	0%	0%	100%
		Language	*Non-RA Index (11)*	42%	40%	18%	0%	0%	100%
		Maths	*Non-RA Index (12)*	48%	38%	14%	1%	0%	100%
			Non-RA Average Index	**66.8%**	**26.4%**	**6.4%**	**0.3%**	**0%**	

Notes: **(i)** In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools (ii) There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test. **(v)** The level of Non-Random assignment depends on the **Sorting Evidence (SoE)** in every school: None if **SoE**=0; Low if **SoE**=1,2 (1,2,3); Medium if **SoE**=3,4 (5,4,6); Med-High if **SoE**=5,6 (7,8,9); and High if **SoE**=7,8 (10,11,12); in brackets conditions for **SoE** categories in Group 2.

Table Results - Cohort 3

Table 8.6: Estimated levels of Non-Random assignment
Group 1 & 2 (Cohort 3)

Non-RA Indexes				Sorting Evidence (SoE)					Total No. of schools
				None	Low	Med	Med-High	High	
Random (RDM)	T-test	GPA	*Non-RA Index (1)*	1,181	260	29	5	0	1,475
		Language	*Non-RA Index (2)*	1,133	277	54	8	3	1,475
		Maths	*Non-RA Index (3)*	1,164	283	23	5	0	1,475
	KS-test	GPA	*Non-RA Index (4)*	1,366	94	12	3	0	1,475
		Language	*Non-RA Index (5)*	1,331	130	12	2	0	1,475
		Maths	*Non-RA Index (6)*	1,353	114	8	0	0	1,475
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*	683	680	106	5	1	1,475
		Language	*Non-RA Index (8)*	860	531	82	1	1	1,475
		Maths	*Non-RA Index (9)*	952	458	62	2	1	1,475
	KS-test	GPA	*Non-RA Index (10)*	736	530	205	4	0	1,475
		Language	*Non-RA Index (11)*	669	540	256	10	0	1,475
		Maths	*Non-RA Index (12)*	728	539	201	7	0	1,475
			Non-RA Average Index	**1,013**	**370**	**88**	**4**	**1**	

Notes: (i) In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools (ii) There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test. **(v)** The level of Non-Random assignment depends on the **Sorting Evidence (SoE)** in every school: None if **SoE**=0; Low if **SoE**=1,2 (1,2,3); Medium if **SoE**=3,4 (5,4,6); Med-High if **SoE**=5,6 (7,8,9); and High if **SoE**=7,8 (10,11,12); in brackets conditions for **SoE** categories in Group 2.

Table 8.7: Estimated levels of Non-Random assignment in percentage
Group 1 & 2 (Cohort 3)

Non-RA Indexes				Sorting Evidence (SoE)					Total schools (%)
				None	Low	Med	Med-High	High	
Random (RDM)	T-test	GPA	*Non-RA Index (1)*	80%	18%	2%	0%	0%	100%
		Language	*Non-RA Index (2)*	77%	19%	4%	1%	0%	100%
		Maths	*Non-RA Index (3)*	79%	19%	2%	0%	0%	100%
	KS-test	GPA	*Non-RA Index (4)*	93%	6%	1%	0%	0%	100%
		Language	*Non-RA Index (5)*	90%	9%	1%	0%	0%	100%
		Maths	*Non-RA Index (6)*	92%	8%	1%	0%	0%	100%
Perfectly Sorted (SRT)	T-test	GPA	*Non-RA Index (7)*	46%	46%	7%	0%	0%	100%
		Language	*Non-RA Index (8)*	58%	36%	6%	0%	0%	100%
		Maths	*Non-RA Index (9)*	65%	31%	4%	0%	0%	100%
	KS-test	GPA	*Non-RA Index (10)*	50%	36%	14%	0%	0%	100%
		Language	*Non-RA Index (11)*	45%	37%	17%	1%	0%	100%
		Maths	*Non-RA Index (12)*	49%	37%	14%	0%	0%	100%
			Non-RA Average Index	**68.7%**	**25.1%**	**5.9%**	**0.3%**	**0%**	

Notes: (i) In total, we have 12 independent Non-Random Assignment measures (from 12 Non-RA Indexes) per group of schools (ii) There are two categories of artificially created counterfactual classes: **Random (RDM)** and **Perfectly Sorted counterfactual (SRT)**. **(iii)** SRT counterfactual can be created based on GPA, Language, or Maths school Marks. **(iv)** To compare real classes with counterfactual classes we apply two statistics test: T-test and KS-test. **(v)** The level of Non-Random assignment depends on the **Sorting Evidence (SoE)** in every school: None if **SoE**=0; Low if **SoE**=1,2 (1,2,3); Medium if **SoE**=3,4 (5,4,6); Med-High if **SoE**=5,6 (7,8,9); and High if **SoE**=7,8 (10,11,12); in brackets conditions for **SoE** categories in Group 2.

8.4 Appendix - Chapter 5

Appendix 5.1 Empirical Bayes kernel distributions

Figure 8.17: Empirical Bayes distribution of Student Ability
4^{th} grade 2005 selected sample cohort

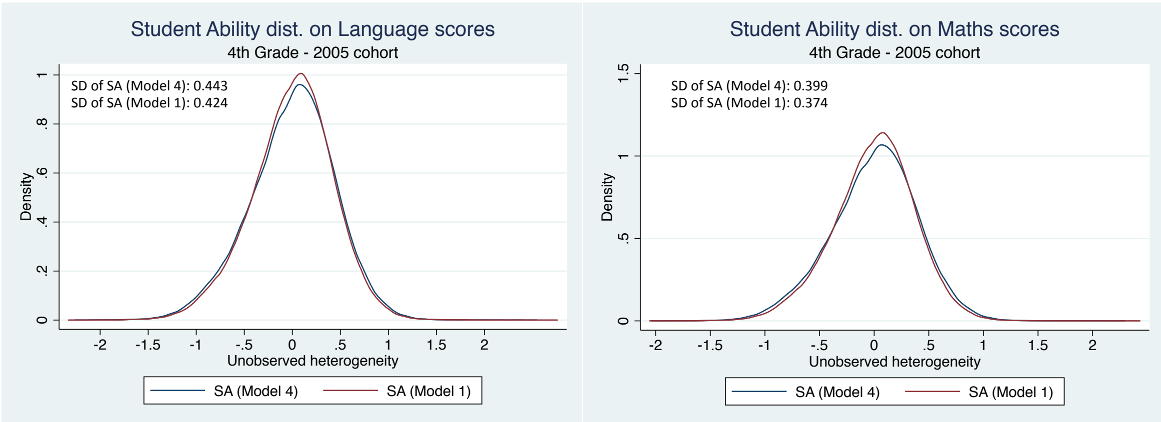


Figure 8.18: Empirical Bayes distribution of Teacher Effects
4^{th} grade 2005 selected sample cohort

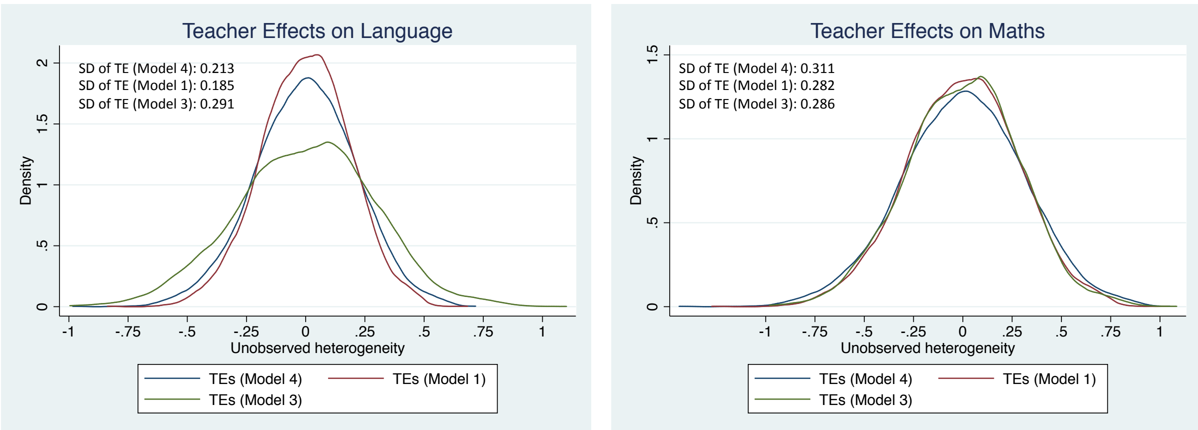
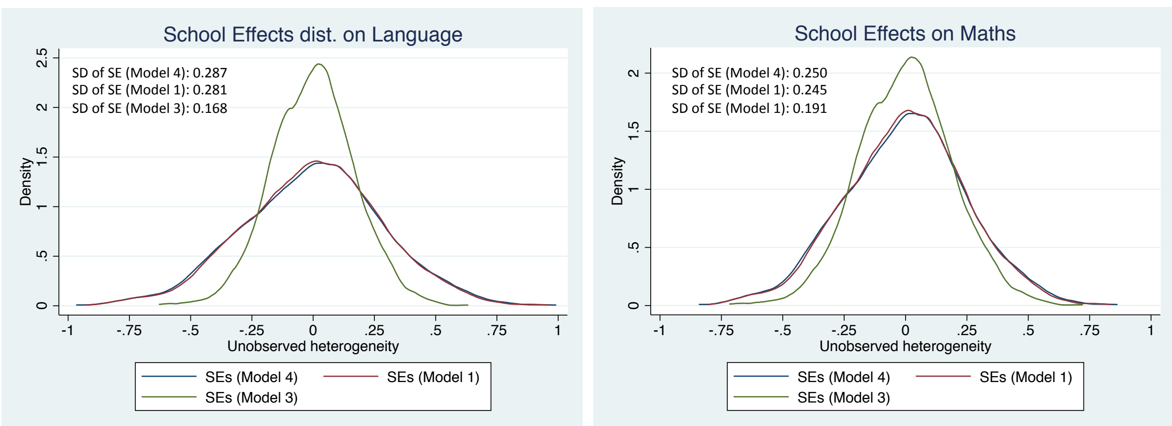


Figure 8.19: Empirical Bayes distribution of School Effects
4^{th} grade 2005 selected sample cohort



Appendix 5.2

Table 8.8: Maximum Likelihood coefficient estimates by gender-type of school 4^{th} grade 2005 selected sample cohort

	Lambda assigned (λ=0.4) **Model 4**				*Lambda assigned (λ=1)* **Model 3**				*Lambda estimated* **Model 1**			
	Girls-only schools		*Boys-only schools*		*Girls-only schools*		*Boys-only schools*		*Girls-only schools*		*Boys-only schools*	
	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)	Language (7)	Maths (8)	Language (9)	Maths (10)	Language (11)	Maths (12)
Loading factors												
School Quality	**1**	1.900	**1**	0.966***	**1**	1.168	**1**	1.312	**1**	0.899***	**1**	0.955
		(1.402)		(0.082)		(0.128)		(0.397)		(0.083)		(0.083)
Teacher Effects	**1**	1.038	**1**	2.353**	**1**	0.925	**1**	0.931	**1**	1.582***	**1**	2.393**
		(0.098)		(0.525)		(0.056)		(0.080)		(0.142)		(0.542)
Individual Ability	1.175***	**1**	1.208***	**1**					1.212***	**1**	1.222***	**1**
	(0.031)		(0.043)						(0.033)		(0.044)	
Student covariates												
Stdzd. Language School Marks 2004									0.488***		0.450***	
									(0.010)		(0.014)	
Stdzd. Maths School Marks 2004										0.511***		0.476***
										(0.009)		(0.013)
Mother education level	0.027***	0.021***	0.026***	0.021***	-0.018***	-0.017**	0.001	0.001	0.022***	0.016***	0.026***	0.019**
	(0.006)	(0.005)	(0.009)	(0.008)	(0.007)	(0.007)	(0.010)	(0.010)	(0.006)	(0.005)	(0.009)	(0.008)
Household income	0.012**	0.018***	0.003	0.012*	-0.024***	-0.012**	-0.008	0.001	0.007	0.014***	0.003	0.011*
	(0.006)	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)	(0.006)	(0.005)	(0.007)	(0.006)
Special needs	-0.557***	-0.545***	-0.036	0.606	0.123	-0.002	-0.082	0.766	-0.485**	-0.486**	-0.039	0.616
	(0.208)	(0.194)	(0.665)	(0.609)	(0.235)	(0.235)	(0.766)	(0.766)	(0.208)	(0.192)	(0.665)	(0.606)
Class and Teacher covariates												
Class size	0.009**	0.010**	0.015**	0.021***	0.006	0.008	-0.004	0.001	-0.001	0.001	0.015**	0.020***
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.008)	(0.008)	(0.005)	(0.005)	(0.006)	(0.007)
Peers average GPA	-0.366***	-0.473***	-0.552***	-0.560***	1.055***	1.029***	1.268***	1.401***	-0.293***	-0.384***	-0.514***	-0.498***
	(0.114)	(0.130)	(0.155)	(0.191)	(0.241)	(0.244)	(0.301)	(0.308)	(0.102)	(0.116)	(0.151)	(0.186)
Gender (Female=1)	-0.115	-0.102	-0.048	-0.050	0.588*	0.716**	0.021	0.018	0.034	0.063	-0.050	-0.050
	(0.205)	(0.222)	(0.056)	(0.073)	(0.301)	(0.294)	(0.118)	(0.116)	(0.168)	(0.206)	(0.055)	(0.072)
Years of experience in the education system	0.000	0.002	0.006***	0.006**	-0.002	0.000	0.001	-0.001	0.002	0.005*	0.006***	0.006**
	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)
Teaching ratio (Hrs teaching / Total hrs contract)	-0.003	-0.004	0.523***	0.432*	-0.118	-0.147	-0.311	-0.473	-0.033	0.002	0.509***	0.406*
	(0.202)	(0.227)	(0.190)	(0.227)	(0.295)	(0.295)	(0.303)	(0.306)	(0.178)	(0.206)	(0.188)	(0.224)
School covariates												
Private voucher school	0.377***	0.341***	0.690***	0.687***	0.302***	0.223**	0.563***	0.545***	0.464***	0.416***	0.688***	0.686***
	(0.061)	(0.077)	(0.145)	(0.151)	(0.100)	(0.104)	(0.137)	(0.141)	(0.074)	(0.078)	(0.143)	(0.148)
Unsubsidised private school	0.860***	1.027***	1.343***	1.319***	0.835***	0.877***	0.567**	0.468**	0.841***	0.993***	1.326***	1.297***
	(0.119)	(0.137)	(0.208)	(0.220)	(0.188)	(0.192)	(0.229)	(0.236)	(0.133)	(0.139)	(0.205)	(0.214)
Rurality (Rural=1)	-0.001	0.085	-0.082	-0.096	-0.148	-0.005	0.322	0.312	-0.047	0.043	-0.071	-0.085
	(0.156)	(0.185)	(0.366)	(0.374)	(0.247)	(0.252)	(0.308)	(0.322)	(0.184)	(0.195)	(0.360)	(0.365)
Constant	1.693**	2.023***	1.968**	2.035**	-6.754***	-6.869***	-7.190***	-7.844***	1.693**	2.023***	1.773**	1.725*
	(0.668)	(0.778)	(0.830)	(1.026)	(1.547)	(1.557)	(1.630)	(1.657)	(0.668)	(0.778)	(0.812)	(0.999)
Prior SD of School Effects: σ_s	0.077	0.147	0.338***	0.326***	0.195***	0.228***	0.141***	0.186***	0.207***	0.186***	0.332***	0.317***
	(0.017)	(0.017)	(0.035)	(0.035)	(0.019)	(0.019)	(0.022)	(0.022)	(0.012)	(0.012)	(0.033)	(0.033)
Prior SD of Teacher Effects: σ_τ	0.241***	0.250***	0.089*	0.210*	0.345***	0.319***	0.375***	0.350***	0.145***	0.229***	0.084***	0.200***
	(0.019)	(0.019)	(0.005)	(0.005)	(0.026)	(0.026)	(0.041)	(0.041)	(0.006)	(0.006)	(0.004)	(0.004)
Prior SD of Individual Ability: σ_α	0.461***	0.392***	0.486***	0.402***					0.455***	0.377***	0.484***	0.396***
	(0.007)	(0.007)	(0.010)	(0.010)					(0.012)	(0.012)	(0.010)	(0.010)
Error variance: σ^2_ε	0.195***	0.195***	0.197***	0.197***	0.525***	0.525***	0.574***	0.574***	0.197***	0.197***	0.198***	0.198***
	(0.004)	(0.004)	(0.006)	(0.006)	(0.008)	(0.008)	(0.012)	(0.012)	(0.004)	(0.004)	(0.006)	(0.006)
Log likelihood		-8952.4		-4601.5		-11328.2		-5927.2		-8914.7		-4600.0
Number of Observations		5,070		2,547		5,070		2,547		5,070		2,547

Notes: **(i)** To estimate the models and obtain the Posterior SD of unobserved heterogeneities, we use the Generalised Structural Equation Modelling (GSEM) programme, available in Stata 13 (StataCorp, 2013). **(ii)** Standard errors in parentheses. **(iii)** *** p<0.001; ** p<0.05; * p<0.01. **(iv)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(v)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600. **(vi)** Both Mother education level and Household income variables are assumed to be the same the year reported in the Simce Parents questionnaire and the year before, for which we do not have records. **(vii)** The base category for the included dummy variables is shown next in brackets: Pupil's gender (Male); Special needs pupils (all the rest pupils); Teacher's gender (Male); *Private voucher* schools (Municipal and Unsubsidised Private school); *Unsubsidised Private* school (Municipal and Private voucher schools); Rurality (Urban area).

Table 8.9: Maximum Likelihood coefficient estimates by type of school dependence
4^{th} grade 2005 selected sample cohort

	Lambda assigned (λ=0.4)				*Lambda assigned (λ=1)*				*Lambda estimated*			
	Model 4				**Model 3**				**Model 1**			
	Municipals and Private Voucher Schools		*Only Municipal Schools*		*Municipals and Private Voucher Schools*		*Only Municipal Schools*		*Municipals and Private Voucher Schools*		*Only Municipal Schools*	
	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)	Language (7)	Maths (8)	Language (9)	Maths (10)	Language (11)	Maths (12)
Loading factors												
School Quality	**1**	0.878***	**1**	0.812***	**1**	1.117***	**1**	1.110**	**1**	0.876***	**1**	0.817***
		(0.015)		(0.024)		(0.034)		(0.050)		(0.014)		(0.023)
Teacher Effects	**1**	1.456***	**1**	1.366***	**1**	0.997	**1**	1.002	**1**	1.508***	**1**	1.429***
		(0.022)		(0.024)		(0.019)		(0.026)		(0.023)		(0.026)
Individual Ability	1.109***	**1**	1.104***	**1**					1.129***	**1**	1.126***	**1**
	(0.006)		(0.007)						(0.006)		(0.008)	
Student covariates												
Stdzd. Language School Marks 2004									0.526***		0.534***	
									(0.003)		(0.003)	
Stdzd. Maths School Marks 2004										0.553***		0.562***
										(0.002)		(0.003)
Gender (Female=1)	0.018***	-0.066***	0.012*	-0.072***	-0.079***	-0.040***	-0.083***	-0.040***	0.006	-0.060***	-0.001	-0.063***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.007)	(0.007)	(0.005)	(0.005)	(0.006)	(0.006)
Mother education level	0.031***	0.028***	0.033***	0.033***	-0.004***	-0.006***	-0.005**	-0.003	0.026***	0.023***	0.028***	0.027***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Household income	0.020***	0.021***	0.022***	0.023***	-0.000	-0.002	-0.003	-0.004	0.018***	0.017***	0.019***	0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Special needs	-0.354***	-0.276***	-0.380***	-0.311***	-0.063	0.012	-0.095	-0.051	-0.313***	-0.225***	-0.338***	-0.258***
	(0.053)	(0.051)	(0.061)	(0.057)	(0.051)	(0.052)	(0.064)	(0.063)	(0.052)	(0.050)	(0.060)	(0.056)
Class and Teacher covariates												
Class size	0.015***	0.016***	0.019***	0.021***	0.009***	0.009***	0.009***	0.009***	0.013***	0.014***	0.016***	0.018***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Peers average GPA	-0.518***	-0.613***	-0.684***	-0.786***	0.618***	0.642***	0.453***	0.505***	-0.346***	-0.399***	-0.413***	-0.454***
	(0.036)	(0.042)	(0.058)	(0.068)	(0.034)	(0.035)	(0.041)	(0.042)	(0.028)	(0.033)	(0.037)	(0.043)
Gender (Female=1)	0.081***	0.069***	0.112***	0.099***	0.048*	0.023	0.048	0.029	0.069***	0.056**	0.091***	0.077**
	(0.020)	(0.024)	(0.030)	(0.037)	(0.025)	(0.025)	(0.033)	(0.034)	(0.018)	(0.023)	(0.027)	(0.032)
Years of experience in the education system	0.004***	0.004***	0.003***	0.003***	0.005***	0.004***	0.005***	0.004***	0.004***	0.004***	0.004***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Teaching ratio (Hrs teaching / Total hrs contract)	-0.056	-0.095*	-0.063	-0.084	-0.023	-0.075	-0.000	-0.035	-0.043	-0.086	-0.045	-0.066
	(0.051)	(0.058)	(0.073)	(0.081)	(0.059)	(0.061)	(0.075)	(0.079)	(0.048)	(0.054)	(0.066)	(0.073)
School covariates												
Private voucher school	0.430***	0.428***			0.377***	0.356***			0.421***	0.416***		
	(0.024)	(0.025)			(0.023)	(0.024)			(0.022)	(0.023)		
Rurality (Rural=1)	0.027	-0.003	0.073	0.038	-0.052	-0.085	-0.040	-0.078	0.015	-0.018	0.051	0.011
	(0.056)	(0.057)	(0.061)	(0.063)	(0.052)	(0.055)	(0.056)	(0.059)	(0.054)	(0.054)	(0.057)	(0.058)
Constant	2.154***	2.784***	2.986***	3.630***	-4.243***	-4.300***	-3.294***	-3.554***	1.773**	1.725*	1.482***	1.769***
	(0.214)	(0.251)	(0.341)	(0.398)	(0.212)	(0.217)	(0.255)	(0.257)	(0.812)	(0.999)	(0.221)	(0.261)
Prior SD of School Effects: σ_s	0.324***	0.285***	0.292***	0.237***	0.239***	0.267***	0.226***	0.251***	0.316***	0.277***	0.285***	0.233***
	(0.006)	(0.006)	(0.008)	(0.008)	(0.005)	(0.005)	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)
Prior SD of Teacher Effects: σ_τ	0.243***	0.354***	0.300***	0.401***	0.339***	0.338***	0.335***	0.335***	0.214***	0.323***	0.251***	0.359***
	(0.004)	(0.004)	(0.009)	(0.009)	(0.006)	(0.006)	(0.007)	(0.007)	(0.003)	(0.003)	(0.005)	(0.005)
Prior SD of Individual Ability: σ_α	0.530***	0.477***	0.542***	0.491***					0.516***	0.457***	0.527***	0.468***
	(0.002)	(0.002)	(0.003)	(0.003)					(0.002)	(0.002)	(0.003)	(0.003)
Error variance: σ_e^2	0.197***	0.197***	0.197***	0.197***	0.551***	0.551***	0.552***	0.552***	0.202***	0.202***	0.202***	0.202***
	(0.006)	(0.006)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Log likelihood		-162590.9		-98325.1		-195828.5		-117220.3		-161781.4		-97746.2
Number of Observations		85,737		51,270		85,737		51,270		85,737		51,270

Notes: **(i)** To estimate the models and obtain the Posterior SD of unobserved heterogeneities, we use the Generalised Structural Equation Modelling (GSEM) programme, available in Stata 13 (StataCorp, 2013). **(ii)** Standard errors in parentheses. **(iii)** *** p<0.001; ** p<0.05; * p<0.01. **(iv)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(v)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600. **(vi)** Both Mother education level and Household income variables are assumed to be the same the year reported in the Simce Parents questionnaire and the year before, for which we do not have records. **(vii)** The base category for the included dummy variables is shown next in brackets: Pupil's gender (Male); Special needs pupils (all the rest pupils); Teacher's gender (Male); *Private voucher* schools (Municipal and Unsubsidised Private school); *Unsubsidised Private* school (Municipal and Private voucher schools); Rurality (Urban area).

Appendix 5.3 Simce Score Distribution

Table 8.10: Simce distribution - Absolute and standardised Simce scores
4^{th} grade 2005 selected sample

Absolute Simce Score

	All schools		Girls-only schools		Boys-only schools	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	138.2	128.3	153.4	145.3	149.9	140.6
5%	164.7	156.6	188.4	179.6	183.9	183.4
10%	183.4	176.3	209.4	198.6	208.9	207.1
25%	222.8	213.0	246.5	231.9	251.7	248.2
50%	263.8	254.2	279.7	267.7	286.6	286.6
75%	296.5	291.3	307.6	299.1	315.9	317.6
90%	323.5	320.1	333.2	324.2	339.6	341.3
95%	338.9	335.5	347.6	338.8	355.2	347.0
99%	362.4	361.7	364.8	363.6	364.8	363.6
75th - 25th	73.7	78.3	61.0	67.2	64.2	69.4
90th - 10th	140.1	143.8	123.8	125.6	130.8	134.2
Mean	258.5	251.1	275.0	264.1	279.9	279.5
SD	**52.3**	**54.2**	**46.9**	**48.1**	**49.6**	**50.8**
Number of students		89,710		5,070		2,547
Number of teachers		3,151		162		86
Number of schools		1,337		66		36

Standardised Simce Score

	All schools		Girls-only schools		Boys-only schools	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-2.207	-2.160	-1.921	-1.853	-1.985	-1.937
5%	-1.708	-1.648	-1.264	-1.232	-1.347	-1.164
10%	-1.357	-1.292	-0.868	-0.888	-0.878	-0.734
25%	-0.616	-0.628	-0.170	-0.286	-0.074	0.009
50%	0.154	0.118	0.452	0.361	0.582	0.703
75%	0.768	0.788	0.976	0.930	1.132	1.264
90%	1.275	1.309	1.459	1.384	1.579	1.693
95%	1.564	1.588	1.729	1.648	1.871	1.796
99%	2.007	2.063	2.051	2.097	2.051	2.097
75th - 25th	1.384	1.417	1.147	1.216	1.206	1.255
90th - 10th	2.632	2.601	2.327	2.272	2.457	2.427
Mean	0.054	0.062	0.364	0.296	0.456	0.574
SD	**0.983**	**0.980**	**0.881**	**0.871**	**0.932**	**0.918**
Number of students		89,710		5,070		2,547
Number of teachers		3,151		162		86
Number of schools		1,337		66		36

Absolute Simce Score

	All schools		Only Municipal schools		Municipal and Priv. Voucher	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	138.2	128.3	134.8	124.9	137.8	127.8
5%	164.7	156.6	158.6	150.0	163.8	155.6
10%	183.4	176.3	175.5	167.5	182.1	174.9
25%	222.8	213.0	210.6	201.1	220.8	211.0
50%	263.8	254.2	251.6	241.3	261.5	251.7
75%	296.5	291.3	284.8	279.1	294.0	288.5
90%	323.5	320.1	311.1	309.0	320.4	317.3
95%	338.9	335.5	327.5	325.4	337.4	332.2
99%	362.4	361.7	358.7	347.5	358.7	361.7
75th - 25th	73.7	78.3	74.2	77.9	73.2	77.5
90th - 10th	140.1	143.8	135.6	141.5	138.2	142.4
Mean	258.5	251.1	247.5	239.8	256.4	248.8
SD	**52.3**	**54.2**	**51.2**	**53.2**	**51.8**	**53.6**
Number of students		1,337		51,270		85,737
Number of teachers		3,151		1,872		2,974
Number of schools		89,710		794		1,271

Standardised Simce Score

	All schools		Only Municipal schools		Municipal and Priv. Voucher	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-2.207	-2.160	-2.270	-2.222	-2.214	-2.169
5%	-1.708	-1.648	-1.822	-1.768	-1.726	-1.666
10%	-1.357	-1.292	-1.506	-1.451	-1.380	-1.317
25%	-0.616	-0.628	-0.846	-0.843	-0.654	-0.664
50%	0.154	0.118	-0.075	-0.116	0.111	0.072
75%	0.768	0.788	0.548	0.568	0.721	0.737
90%	1.275	1.309	1.042	1.109	1.217	1.258
95%	1.564	1.588	1.351	1.405	1.536	1.528
99%	2.007	2.063	1.937	1.805	1.937	2.063
75th - 25th	1.384	1.417	1.394	1.410	1.375	1.401
90th - 10th	2.632	2.601	2.548	2.559	2.597	2.576
Mean	0.054	0.062	-0.152	-0.143	0.014	0.019
SD	**0.983**	**0.980**	**0.963**	**0.962**	**0.974**	**0.969**
Number of students		1,337		51,270		85,737
Number of teachers		3,151		1,872		2,974
Number of schools		89,710		794		1,271

Appendix 5.4 Empirical Bayes distribution tables

Table 8.11: Empirical Bayes distributions - All schools
4^{th} grade 2005 selected sample

Distribution of Teacher Effects

	Lambda = 1		Lambda reported (lambda = 0.442)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.685	-0.674	-0.512	-0.750	-0.444	-0.675
5%	-0.495	-0.488	-0.356	-0.521	-0.307	-0.467
10%	-0.378	-0.372	-0.270	-0.395	-0.236	-0.359
25%	-0.194	-0.192	-0.142	-0.208	-0.126	-0.191
50%	0.008	0.008	0.003	0.004	0.005	0.007
75%	0.192	0.189	0.145	0.212	0.125	0.191
90%	0.360	0.355	0.267	0.391	0.234	0.356
95%	0.455	0.449	0.343	0.501	0.298	0.453
99%	0.696	0.686	0.500	0.732	0.437	0.664
75th - 25th	0.386	0.380	0.287	0.420	0.251	0.382
90th - 10th	0.738	0.727	0.537	0.786	0.470	0.715
Mean of TEs	0.000	0.000	0.000	0.000	0.000	0.000
SD of (TE)	**0.291**	**0.286**	**0.213**	**0.311**	**0.185**	**0.282**
Number of students						89,710
Number of teachers						3,151
Number of schools						1,337

Distribution of School Effects

	Lambda = 1		Lambda reported (lambda = 0.442)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.399	-0.456	-0.711	-0.618	-0.708	-0.615
5%	-0.274	-0.313	-0.470	-0.409	-0.457	-0.397
10%	-0.208	-0.238	-0.381	-0.332	-0.364	-0.316
25%	-0.115	-0.131	-0.192	-0.167	-0.187	-0.162
50%	0.003	0.004	0.008	0.007	0.009	0.008
75%	0.110	0.126	0.190	0.166	0.186	0.162
90%	0.208	0.238	0.359	0.312	0.354	0.308
95%	0.279	0.318	0.474	0.412	0.466	0.405
99%	0.394	0.450	0.672	0.585	0.655	0.569
75th - 25th	0.225	0.257	0.382	0.332	0.373	0.324
90th - 10th	0.417	0.475	0.740	0.644	0.718	0.624
Mean of SEs	0.000	0.000	0.000	0.000	0.000	0.000
SD of (SE)	**0.168**	**0.191**	**0.287**	**0.250**	**0.281**	**0.245**
Number of students						89,710
Number of teachers						3,151
Number of schools						1,337

Distribution of Student Ability

	Lambda = 1		Lambda reported (lambda = 0.442)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%			-1.112	-1.001	-1.072	-0.946
5%			-0.779	-0.702	-0.743	-0.655
10%			-0.585	-0.526	-0.555	-0.490
25%			-0.278	-0.250	-0.264	-0.233
50%			0.029	0.026	0.024	0.022
75%			0.300	0.270	0.283	0.250
90%			0.539	0.485	0.516	0.455
95%			0.687	0.619	0.659	0.581
99%			0.964	0.868	0.936	0.825
75th - 25th			0.577	0.520	0.547	0.483
90th - 10th			1.124	1.011	1.071	0.944
Mean of SA			0.000	0.000	0.000	0.000
SD of (SA)			**0.443**	**0.399**	**0.424**	**0.374**
Number of students						89,710
Number of teachers						3,151
Number of schools						1,337

Table 8.12: Empirical Bayes distributions - Girls-only schools 4^{th} grade 2005 selected sample

Distribution of Teacher Effects - Girls only

	Lambda = 1		**Lambda reported (lambda = 0.442)**		**Lambda estimaded**	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.709	-0.656	-0.470	-0.488	-0.276	-0.437
5%	-0.452	-0.419	-0.335	-0.347	-0.193	-0.306
10%	-0.373	-0.345	-0.291	-0.302	-0.156	-0.246
25%	-0.199	-0.184	-0.130	-0.135	-0.077	-0.121
50%	-0.008	-0.008	0.006	0.006	-0.008	-0.012
75%	0.213	0.197	0.126	0.131	0.087	0.137
90%	0.391	0.362	0.270	0.281	0.161	0.255
95%	0.527	0.488	0.366	0.380	0.193	0.305
99%	0.719	0.665	0.564	0.585	0.267	0.422
75th - 25th	0.412	0.381	0.256	0.265	0.163	0.258
90th - 10th	0.764	0.707	0.562	0.583	0.317	0.501
Mean of TEs	0.000	0.000	0.000	0.000	0.000	0.000
SD of (TE)	**0.303**	**0.280**	**0.215**	**0.223**	**0.123**	**0.194**
Number of students						5,070
Number of teachers						162
Number of schools						66

Distribution of School Effects - Girls only

	Lambda = 1		**Lambda reported (lambda = 0.442)**		**Lambda estimaded**	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.359	-0.419	-0.159	-0.301	-0.407	-0.366
5%	-0.195	-0.228	-0.098	-0.186	-0.264	-0.238
10%	-0.156	-0.182	-0.074	-0.141	-0.220	-0.197
25%	-0.101	-0.118	-0.040	-0.076	-0.097	-0.087
50%	-0.003	-0.004	-0.005	-0.010	-0.010	-0.009
75%	0.122	0.142	0.045	0.085	0.080	0.072
90%	0.189	0.220	0.083	0.158	0.254	0.229
95%	0.237	0.276	0.094	0.180	0.286	0.257
99%	0.357	0.417	0.150	0.285	0.531	0.478
75th - 25th	0.223	0.260	0.085	0.161	0.177	0.159
90th - 10th	0.344	0.402	0.157	0.299	0.474	0.426
Mean of SEs	0.008	0.009	0.000	0.000	0.000	0.000
SD of (SE)	**0.142**	**0.166**	**0.061**	**0.115**	**0.182**	**0.164**
Number of students						5,070
Number of teachers						162
Number of schools						66

Distribution of Student Ability - Girls only

	Lambda = 1		**Lambda reported (lambda = 0.442)**		**Lambda estimaded**	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%			-0.968	-0.824	-0.963	-0.794
5%			-0.655	-0.557	-0.636	-0.525
10%			-0.497	-0.423	-0.478	-0.395
25%			-0.217	-0.185	-0.214	-0.176
50%			0.028	0.024	0.027	0.022
75%			0.241	0.205	0.236	0.195
90%			0.437	0.372	0.433	0.358
95%			0.563	0.479	0.546	0.450
99%			0.788	0.671	0.785	0.647
75th - 25th			0.458	0.390	0.450	0.371
90th - 10th			0.934	0.795	0.912	0.752
Mean of SA			0.000	0.000	0.000	0.000
SD of (SA)			**0.367**	**0.313**	**0.360**	**0.297**
Number of students						5,070
Number of teachers						162
Number of schools						66

Table 8.13: Empirical Bayes distributions - Boys-only schools 4^{th} grade 2005 selected sample

Distribution of Teacher Effects - Boys only

	Lambda = 1		Lambda reported (lambda = 0.442)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.698	-0.650	-0.202	-0.475	-0.197	-0.472
5%	-0.540	-0.502	-0.117	-0.275	-0.118	-0.283
10%	-0.408	-0.380	-0.100	-0.234	-0.098	-0.235
25%	-0.252	-0.235	-0.036	-0.085	-0.038	-0.091
50%	0.049	0.046	-0.004	-0.010	-0.002	-0.004
75%	0.224	0.209	0.051	0.120	0.049	0.117
90%	0.372	0.346	0.092	0.217	0.091	0.218
95%	0.514	0.478	0.114	0.268	0.112	0.268
99%	0.817	0.761	0.207	0.486	0.206	0.493
75th - 25th	0.477	0.444	0.087	0.205	0.087	0.208
90th - 10th	0.780	0.726	0.192	0.451	0.189	0.453
Mean of TEs	0.000	0.000	0.000	0.000	0.000	0.000
SD of (TE)	**0.340**	**0.316**	**0.072**	**0.170**	**0.071**	**0.170**
Number of students						2,547
Number of teachers						86
Number of schools						36

Distribution of School Effects - Boys only

	Lambda = 1		Lambda reported (lambda = 0.442)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.254	-0.334	-0.769	-0.742	-0.752	-0.718
5%	-0.182	-0.239	-0.439	-0.425	-0.440	-0.420
10%	-0.111	-0.145	-0.362	-0.350	-0.361	-0.345
25%	-0.052	-0.068	-0.269	-0.260	-0.264	-0.252
50%	0.004	0.005	0.013	0.013	0.013	0.012
75%	0.073	0.096	0.250	0.242	0.247	0.236
90%	0.105	0.138	0.456	0.441	0.446	0.426
95%	0.125	0.164	0.567	0.548	0.553	0.528
99%	0.144	0.189	0.685	0.662	0.668	0.638
75th - 25th	0.125	0.164	0.519	0.502	0.512	0.489
90th - 10th	0.216	0.283	0.819	0.791	0.807	0.771
Mean of SEs	0.004	0.005	0.000	0.000	0.000	0.000
SD of (SE)	**0.092**	**0.121**	**0.329**	**0.317**	**0.323**	**0.308**
Number of students						2,547
Number of teachers						86
Number of schools						36

Distribution of Student Ability - Boys only

	Lambda = 1		Lambda reported (lambda = 0.442)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%			-1.293	-1.069	-1.107	-0.906
5%			-0.837	-0.692	-0.724	-0.593
10%			-0.619	-0.511	-0.502	-0.411
25%			-0.252	-0.208	-0.216	-0.177
50%			0.046	0.038	0.035	0.029
75%			0.308	0.254	0.257	0.211
90%			0.521	0.431	0.462	0.378
95%			0.658	0.544	0.586	0.479
99%			0.965	0.797	0.816	0.668
75th - 25th			0.560	0.463	0.473	0.387
90th - 10th			1.140	0.942	0.964	0.789
Mean of SA			0.000	0.000	0.000	0.000
SD of (SA)			**0.456**	**0.377**	**0.390**	**0.320**
Number of students						2,547
Number of teachers						86
Number of schools						36

Table 8.14: Empirical Bayes distributions - Municipal & Private Voucher schools 4^{th} grade 2005 selected sample

Distribution of Teacher Effects - Municipal and Private Voucher Schools

	Lambda = 1		Lambda reported (lambda = 0.458)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.678	-0.676	-0.522	-0.760	-0.453	-0.683
5%	-0.486	-0.485	-0.360	-0.525	-0.314	-0.474
10%	-0.371	-0.369	-0.274	-0.399	-0.244	-0.368
25%	-0.191	-0.191	-0.145	-0.211	-0.129	-0.194
50%	0.010	0.010	0.002	0.003	0.005	0.007
75%	0.192	0.192	0.146	0.213	0.129	0.195
90%	0.352	0.351	0.273	0.397	0.239	0.361
95%	0.444	0.443	0.348	0.506	0.305	0.460
99%	0.678	0.676	0.506	0.736	0.445	0.671
75th - 25th	0.384	0.383	0.291	0.424	0.258	0.390
90th - 10th	0.723	0.721	0.547	0.796	0.484	0.729
Mean of TEs	0.000	0.000	0.000	0.000	0.000	0.000
SD of (TE)	**0.286**	**0.285**	**0.215**	**0.314**	**0.190**	**0.286**
Number of students						85,737
Number of teachers						2,974
Number of schools						1,271

Distribution of School Effects - Municipal and Private Voucher Schools

	Lambda = 1		Lambda reported (lambda = 0.458)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.398	-0.445	-0.720	-0.632	-0.712	-0.624
5%	-0.289	-0.323	-0.477	-0.419	-0.468	-0.410
10%	-0.209	-0.233	-0.386	-0.339	-0.370	-0.324
25%	-0.118	-0.131	-0.202	-0.177	-0.193	-0.169
50%	0.007	0.008	0.008	0.007	0.009	0.008
75%	0.126	0.141	0.193	0.169	0.189	0.165
90%	0.239	0.267	0.372	0.327	0.360	0.315
95%	0.304	0.339	0.486	0.426	0.471	0.413
99%	0.431	0.482	0.678	0.595	0.659	0.577
75th - 25th	0.244	0.272	0.394	0.346	0.382	0.335
90th - 10th	0.448	0.500	0.759	0.666	0.730	0.640
Mean of SEs	0.007	0.008	0.000	0.000	0.000	0.000
SD of (SE)	**0.179**	**0.199**	**0.293**	**0.257**	**0.287**	**0.251**
Number of students						85,737
Number of teachers						2,974
Number of schools						1,271

Distribution of Student Ability - Municipal and Private Voucher Schools

	Lambda = 1		Lambda reported (lambda = 0.458)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%			-1.106	-0.998	-1.072	-0.949
5%			-0.777	-0.700	-0.745	-0.660
10%			-0.586	-0.529	-0.559	-0.495
25%			-0.280	-0.253	-0.267	-0.237
50%			0.028	0.025	0.024	0.022
75%			0.301	0.272	0.286	0.254
90%			0.542	0.488	0.520	0.460
95%			0.689	0.622	0.663	0.587
99%			0.963	0.869	0.939	0.832
75th - 25th			0.581	0.524	0.554	0.490
90th - 10th			1.128	1.017	1.079	0.956
Mean of SA			0.000	0.000	0.000	0.000
SD of (SA)			**0.443**	**0.400**	**0.426**	**0.378**
Number of students						85,737
Number of teachers						2,974
Number of schools						1,271

Table 8.15: Empirical Bayes distributions - Municipal schools 4^{th} grade 2005 selected sample

Distribution of Teacher Effects - Municipal Schools

	Lambda = 1		Lambda reported (lambda = 0.455)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.665	-0.666	-0.643	-0.878	-0.521	-0.745
5%	-0.471	-0.472	-0.467	-0.638	-0.382	-0.546
10%	-0.370	-0.370	-0.362	-0.494	-0.304	-0.434
25%	-0.202	-0.203	-0.184	-0.251	-0.153	-0.218
50%	0.004	0.004	0.003	0.004	0.005	0.007
75%	0.197	0.198	0.188	0.256	0.156	0.222
90%	0.351	0.351	0.347	0.475	0.284	0.406
95%	0.443	0.444	0.454	0.620	0.375	0.536
99%	0.692	0.694	0.636	0.868	0.515	0.736
75th - 25th	0.399	0.400	0.371	0.507	0.308	0.440
90th - 10th	0.720	0.722	0.709	0.969	0.588	0.840
Mean of TEs	0.000	0.000	0.000	0.000	0.000	0.000
SD of (TE)	**0.286**	**0.286**	**0.275**	**0.376**	**0.225**	**0.322**
Number of students						51,270
Number of teachers						1,872
Number of schools						794

Distribution of School Effects - Municipal Schools

	Lambda = 1		Lambda reported (lambda = 0.455)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.376	-0.417	-0.576	-0.468	-0.542	-0.443
5%	-0.259	-0.287	-0.412	-0.335	-0.410	-0.335
10%	-0.196	-0.218	-0.332	-0.269	-0.320	-0.261
25%	-0.107	-0.118	-0.175	-0.142	-0.178	-0.146
50%	0.002	0.002	0.002	0.002	0.000	0.000
75%	0.116	0.129	0.161	0.131	0.158	0.129
90%	0.224	0.248	0.315	0.256	0.323	0.264
95%	0.288	0.320	0.422	0.343	0.418	0.342
99%	0.420	0.466	0.642	0.521	0.659	0.539
75th - 25th	0.223	0.247	0.336	0.273	0.336	0.275
90th - 10th	0.420	0.466	0.647	0.525	0.643	0.525
Mean of SEs	0.007	0.007	0.000	0.000	0.000	0.000
SD of (SE)	**0.168**	**0.186**	**0.255**	**0.207**	**0.252**	**0.206**
Number of students						51,270
Number of teachers						1,872
Number of schools						794

Distribution of Student Ability - Municipal Schools

	Lambda = 1		Lambda reported (lambda = 0.455)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)	Language (5)	Maths (6)
1%			-1.094	-0.991	-1.058	-0.939
5%			-0.782	-0.708	-0.750	-0.666
10%			-0.601	-0.545	-0.571	-0.507
25%			-0.300	-0.272	-0.284	-0.252
50%			0.022	0.020	0.020	0.018
75%			0.311	0.282	0.294	0.262
90%			0.566	0.513	0.540	0.479
95%			0.720	0.653	0.689	0.612
99%			1.008	0.913	0.980	0.871
75th - 25th			0.612	0.554	0.579	0.514
90th - 10th			1.168	1.058	1.111	0.987
Mean of SA			0.000	0.000	0.000	0.000
SD of (SA)			**0.456**	**0.413**	**0.437**	**0.388**
Number of students						51,270
Number of teachers						1,872
Number of schools						794

8.5 Appendix - Chapter 6

Appendix 6.1 Summary statistics 4^{th} grade cohorts (2005-2009)

Table 8.16: 4^{th} grade cohort 2005 - Before and after selection

Before selection	Mean	Std. Dev.	Min	Max	After selection	Mean	Std. Dev.	Min	Max
Pupil Level					**Pupil Level**				
GPA 4th 2005	5.7	0.9	1	7	GPA 4th 2005	5.9	0.6	1	7
GPA 3th 2004	5.9	0.8	1	7	GPA 3th 2004	6.0	0.8	1	7
School Language Marks 2005	5.4	0.8	1	7	School Language Marks 2005	5.5	0.8	1	7
School Maths Marks 2005	5.3	0.9	1	7	School Maths Marks 2005	5.4	0.9	1	7
School Language Marks 2004	5.6	0.8	1	7	School Language Marks 2004	5.7	0.8	1	7
School Maths Marks 2004	5.5	0.9	1	7	School Maths Marks 2004	5.6	0.9	1	7
Language Simce Scores 4th 2005	255.6	53.2	103.1	397.2	Language Simce Scores 4th 2005	258.5	52.3	103.1	364.8
Maths Simce Scores 4th 2005	247.7	55.3	90.7	363.6	Maths Simce Scores 4th 2005	251.1	54.1	91.7	363.6
Stdsed. Lang. Simce Score 4th 2005	0.0	1.0	-2.9	2.7	Stdsed. Lang. Simce Score 4th 2005	0.1	0.98	-2.9	2.1
Stdsed. Maths Simce Score 4th 2005	0.0	1.0	-2.8	2.1	Stdsed. Maths Simce Score 4th 2005	0.1	0.98	-2.8	2.1
Gender (Female=1)	0.5	0.5	0.0	1.0	Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.2	0.6	8	12	Age	9.1	0.4	8	12
Attendance	92.5	14.0	0	100	Attendance	94.4	6.0	0	100
Mother Education	2.8	2.1	0	8	Mother Education	2.9	2.0	0	8
Father Education	3.1	2.4	0	8	Father Education	3.1	2.2	0	8
Household Income	2.2	2.7	0	12	Household Income	2.2	2.6	0	12
Class Level					**Class Level**				
Class size	23.3	14.6	1	54	Class size	28.4	6.8	15	45
Peers Average GPA	5.7	0.5	1	7	Peers Average GPA	5.7	0.4	2.342	6.688
Teacher Level					**Teacher Level**				
Gender (Female=1)	0.83	0.4	0	1	Gender (Female=1)	0.89	0.3	0	1
Years of experience in the system	22.8	13.2	0	40	Years of experience in the system	20.0	11.5	0	40
(Teaching hrs / Contract hrs) Ratio	0.9	0.2	0.02	1	(Teaching hrs / Contract hrs) Ratio	0.9	0.1	0.11	1
School Level					**School Level**				
Municipal Schools	0.60	0.5	0	1	Municipal Schools	0.59	0.5	0	1
Private Voucher Schools	0.35	0.5	0	1	Private Voucher Schools	0.36	0.5	0	1
Unsubsidised Private Schools	0.05	0.2	0	1	Unsubsidised Private Schools	0.05	0.2	0	1
Rural Area	0.49	0.5	0	1	Rural Area	0.04	0.2	0	1
Number of classes per grade	1.38	0.8	1	12	Number of classes per grade	2.36	0.6	2	5
Number of students per grade	32.16	37.6	1	535	Number of students per grade	66.98	25.2	30	209
Number of students	268,162					89,278			
Number of teachers	12,233					3,140			
Number of schools	8,338					1,333			

Notes: (i) The selection sample was made for identification, and the most important sources of dropped observations were schools with less than two classes per grade and schools and schools with specialised teachers only, both sum up to 20% of the original pupil's observations in the cohort. **(ii)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(iii)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600.

Figure 8.20: Kernel distributions: Stdsed. Simce Scores - Before and After Selection
4^{th} grade cohort 2005

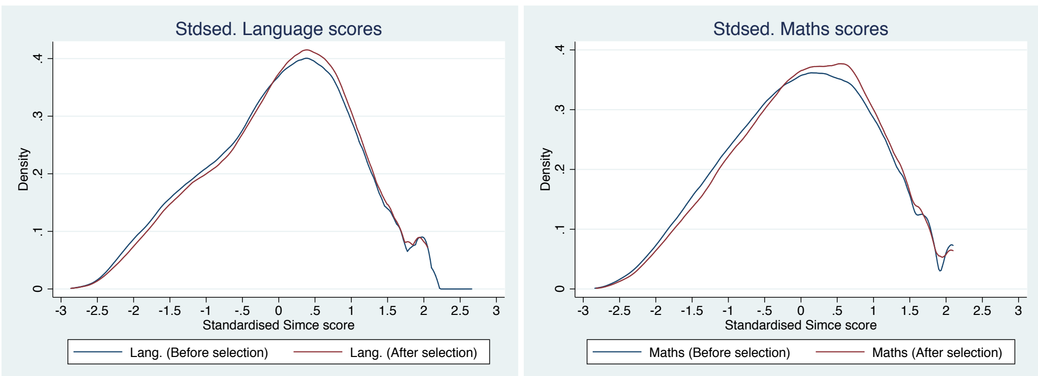


Table 8.17: 4^{th} grade cohort 2006 - Before and after selection

Before selection	Mean	Std. Dev.	Min	Max	After selection	Mean	Std. Dev.	Min	Max
Pupil Level					**Pupil Level**				
GPA 4th 2006	5.7	0.9	1	7	GPA 4th 2006	5.9	0.6	1	7
GPA 3th 2005	5.9	0.8	1	7	GPA 3th 2005	6.0	0.7	1	7
School Language Marks 2006	5.4	0.8	1	7	School Language Marks 2006	5.5	0.8	1.65	7
School Maths Marks 2006	5.3	0.9	1	7	School Maths Marks 2006	5.4	0.9	1.7	7
School Language Marks 2005	5.6	0.8	1	7	School Language Marks 2005	5.7	0.8	1	7
School Maths Marks 2005	5.5	0.9	1	7	School Maths Marks 2005	5.6	0.8	1	7
Language Simce Scores 4th 2006	253.4	53.8	101.9	373.4	Language Simce Scores 4th 2006	255.7	53.3	102.7	373.4
Maths Simce Scores 4th 2006	247.8	55.8	74.3	359.4	Maths Simce Scores 4th 2006	251.6	54.9	76.6	359.4
Stdsed. Lang. Simce Score 4th 2006	0.0	1.0	-2.8	2.2	Stdsed. Lang. Simce Score 4th 2006	0.0	0.99	-2.8	2.2
Stdsed. Maths Simce Score 4th 2006	0.0	1.0	-3.1	2.0	Stdsed. Maths Simce Score 4th 2006	0.1	0.98	-3.1	2.0
Gender (Female=1)	0.5	0.5	0.0	1.0	Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.3	0.6	8	12	Age	9.2	0.5	8	12
Attendance	92.4	14.4	0	100	Attendance	94.3	6.3	0	100
Mother Education	2.9	1.9	0	8	Mother Education	3.0	1.8	0	8
Father Education	2.9	2.0	0	8	Father Education	3.0	2.0	0	8
Household Income	3.4	2.8	0	12	Household Income	3.4	2.7	0	12
Class Level					**Class Level**				
Class size	23.5	14.6	1	57	Class size	29.3	6.8	15	45
Peers Average GPA	5.7	0.5	1	6.9	Peers Average GPA	5.7	0.4	3.767	6.73
Teacher Level					**Teacher Level**				
Gender (Female=1)	0.83	0.4	0	1	Gender (Female=1)	0.90	0.3	0	1
Years of experience in the system	22.8	13.2	0	40	Years of experience in the system	20.4	11.7	0	40
(Teaching hrs / Contract hrs) Ratio	0.9	0.2	0.02	1	(Teaching hrs / Contract hrs) Ratio	0.9	0.1	0.08	1
School Level					**School Level**				
Municipal Schools	0.59	0.5	0	1	Municipal Schools	0.58	0.5	0	1
Private Voucher Schools	0.36	0.5	0	1	Private Voucher Schools	0.36	0.5	0	1
Unsubsidised Private Schools	0.05	0.2	0	1	Unsubsidised Private Schools	0.06	0.2	0	1
Rural Area	0.48	0.5	0	1	Rural Area	0.04	0.2	0	1
Number of classes per grade	1.37	0.7	1	12	Number of classes per grade	2.32	0.6	2	5
Number of students per grade	32.14	37.0	1	511	Number of students per grade	68.03	26.1	31	206
Number of students	265,681					89,261			
Number of teachers	12,166					3,045			
Number of schools	8,267					1,312			

Notes: **(i)** The selection sample was made for identification, and the most important sources of dropped observations were schools with less than two classes per grade and schools and schools with specialised teachers only, both sum up to 20% of the original pupil's observations in the cohort. **(ii)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(iii)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600.

Figure 8.21: Kernel distributions: Stdsed. Simce Scores - Before and After Selection
4^{th} grade cohort 2006

Table 8.18: 4^{th} grade cohort 2007 - Before and after selection

Before selection	Mean	Std. Dev.	Min	Max	After selection	Mean	Std. Dev.	Min	Max
Pupil Level					**Pupil Level**				
GPA 4th 2007	5.7	0.9	1	7	GPA 4th 2007	5.9	0.7	1	7
GPA 3th 2006	5.9	0.8	1	7	GPA 3th 2006	5.9	0.8	1	7
School Language Marks 2007	5.4	0.8	1	7	School Language Marks 2007	5.5	0.8	1	7
School Maths Marks 2007	5.3	0.9	1	7	School Maths Marks 2007	5.4	0.9	1	7
School Language Marks 2006	5.5	0.8	1	7	School Language Marks 2006	5.6	0.8	1	7
School Maths Marks 2006	5.5	0.9	1	7	School Maths Marks 2006	5.5	0.8	1	7
Language Simce Scores 4th 2007	254.7	53.5	109.6	379.4	Language Simce Scores 4th 2007	257.4	52.9	112	379.4
Maths Simce Scores 4th 2007	246.0	56.4	87.1	369.6	Maths Simce Scores 4th 2007	249.6	55.6	88.4	369.6
Stdsed. Lang. Simce Score 4th 2007	0.0	1.0	-2.8	2.2	Stdsed. Lang. Simce Score 4th 2007	0.1	0.99	-2.7	2.3
Stdsed. Maths Simce Score 4th 2007	0.0	1.0	-3.1	2.0	Stdsed. Maths Simce Score 4th 2007	0.1	0.99	-2.8	2.2
Gender (Female=1)	0.5	0.5	0.0	1.0	Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.3	0.6	8	12	Age	9.2	0.4	8	12
Attendance	93.9	8.5	0	100	Attendance	94.2	7.9	0	100
Mother Education	3.0	2.0	0	8	Mother Education	3.0	1.9	0	8
Father Education	2.9	1.9	0	8	Father Education	3.0	1.8	0	8
Household Income	3.6	2.8	0	12	Household Income	3.6	2.7	0	12
Class Level					**Class Level**				
Class size	23.2	14.4	1	50	Class size	29.1	6.8	15	45
Peers Average GPA	5.7	0.5	1	7	Peers Average GPA	5.7	0.3	4.541	6.685
Teacher Level					**Teacher Level**				
Gender (Female=1)	0.84	0.4	0	1	Gender (Female=1)	0.90	0.3	0	1
Years of experience in the system	22.7	13.4	0	40	Years of experience in the system	20.6	12.0	0	40
(Teaching hrs / Contract hrs) Ratio	0.9	0.2	0.02	1	(Teaching hrs / Contract hrs) Ratio	0.9	0.1	0.05	1
School Level					**School Level**				
Municipal Schools	0.58	0.5	0	1	Municipal Schools	0.58	0.5	0	1
Private Voucher Schools	0.37	0.5	0	1	Private Voucher Schools	0.36	0.5	0	1
Unsubsidised Private Schools	0.05	0.2	0	1	Unsubsidised Private Schools	0.06	0.2	0	1
Rural Area	0.47	0.5	0	1	Rural Area	0.03	0.2	0	1
Number of classes per grade	1.35	0.7	1	11	Number of classes per grade	2.30	0.6	2	5
Number of students per grade	31.46	35.8	1	473	Number of students per grade	66.85	25.2	32	193
Number of students	257,344					63,578			
Number of teachers	12,052					2,186			
Number of schools	8,180					951			

Notes: **(i)** The selection sample was made for identification, and the most important sources of dropped observations were schools with less than two classes per grade and schools and schools with specialised teachers only, both sum up to 20% of the original pupil's observations in the cohort. **(ii)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(iii)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600.

Figure 8.22: Kernel distributions: Stdsed. Simce Scores - Before and After Selection
4^{th} grade cohort 2007

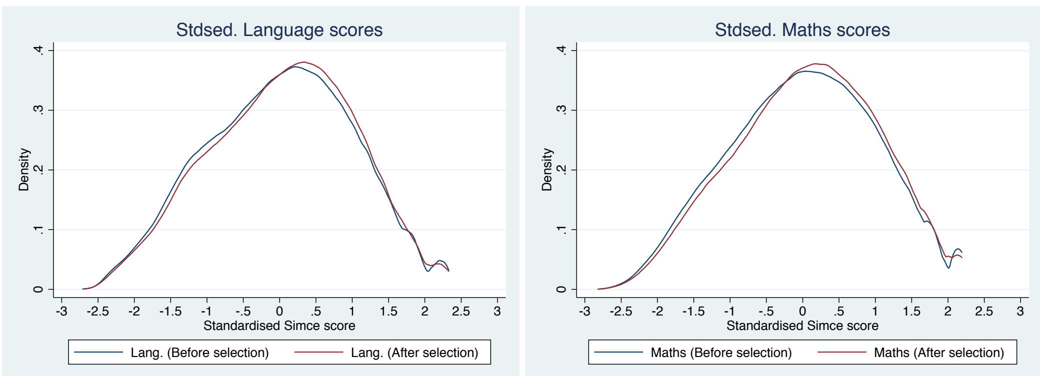


Table 8.19: 4^{th} grade cohort 2008 - Before and after selection

Before selection	**Mean**	**Std. Dev.**	**Min**	**Max**	**After selection**	**Mean**	**Std. Dev.**	**Min**	**Max**
Pupil Level					**Pupil Level**				
GPA 4th 2008	5.7	0.9	1	7	GPA 4th 2008	5.9	0.7	1	7
GPA 3th 2007	5.8	0.8	1	7	GPA 3th 2007	5.9	0.8	1	7
School Language Marks 2008	5.4	0.8	1	7	School Language Marks 2008	5.5	0.8	1	7
School Maths Marks 2008	5.3	0.9	1	7	School Maths Marks 2008	5.3	0.9	1	7
School Language Marks 2007	5.5	0.8	1	7	School Language Marks 2007	5.6	0.8	1	7
School Maths Marks 2007	5.4	0.9	1	7	School Maths Marks 2007	5.5	0.8	1	7
Language Simce Scores 4th 2008	260.6	53.6	124.0	382.5	Language Simce Scores 4th 2008	264.0	53.20	124.9	382.5
Maths Simce Scores 4th 2008	247.3	55.0	101.3	377.5	Maths Simce Scores 4th 2008	251.9	54.38	103.5	377.5
Stdsed. Lang. Simce Score 4th 2008	0.0	1.0	-2.548	2.273	Stdsed. Lang. Simce Score 4th 2008	0.1	1.0	-2.531	2.273
Stdsed. Maths Simce Score 4th 2008	0.0	1.0	-2.7	2.4	Stdsed. Maths Simce Score 4th 2008	0.1	1.0	-2.6	2.4
Gender (Female=1)	0.5	0.5	0.0	1.0	Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.3	0.6	8	12	Age	9.2	0.4	8	12
Attendance	91.8	13.9	0	100	Attendance	93.1	9.7	0	100
Mother Education	2.9	2.0	0	8	Mother Education	3.0	1.9	0	8
Father Education	3.0	2.1	0	8	Father Education	3.1	2.0	0	8
Household Income	2.8	3.0	0	12	Household Income	2.9	2.9	0	12
Class Level					**Class Level**				
Class size	23.1	14.3	1	49	Class size	28.5	6.9	15	45
Peers Average GPA	5.7	0.5	1	7	Peers Average GPA	5.7	0.3	4.512	6.688
Teacher Level					**Teacher Level**				
Gender (Female=1)	0.83	0.4	0	1	Gender (Female=1)	0.89	0.3	0	1
Years of experience in the system	21.8	13.8	0	40	Years of experience in the system	18.8	12.0	0	40
(Teaching hrs / Contract hrs) Ratio	0.9	0.2	0.03	1	(Teaching hrs / Contract hrs) Ratio	0.9	0.1	0.05	1
School Level					**School Level**				
Municipal Schools	0.57	0.5	0	1	Municipal Schools	0.55	0.5	0	1
Private Voucher Schools	0.37	0.5	0	1	Private Voucher Schools	0.38	0.5	0	1
Unsubsidised Private Schools	0.05	0.2	0	1	Unsubsidised Private Schools	0.06	0.2	0	1
Rural Area	0.46	0.5	0	1	Rural Area	0.04	0.2	0	1
Number of classes per grade	1.35	0.7	1	11	Number of classes per grade	2.30	0.6	2	5
Number of students per grade	31.16	35.1	1	469	Number of students per grade	65.74	25.4	30	204
Number of students	254,673					73,825			
Number of teachers	12,092					2,587			
Number of schools	8,174					1,123			

Notes: (i) The selection sample was made for identification, and the most important sources of dropped observations were schools with less than two classes per grade and schools and schools with specialised teachers only, both sum up to 20% of the original pupil's observations in the cohort. **(ii)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(iii)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600.

Figure 8.23: Kernel distributions: Stdsed. Simce Scores - Before and After Selection
4^{th} grade cohort 2008

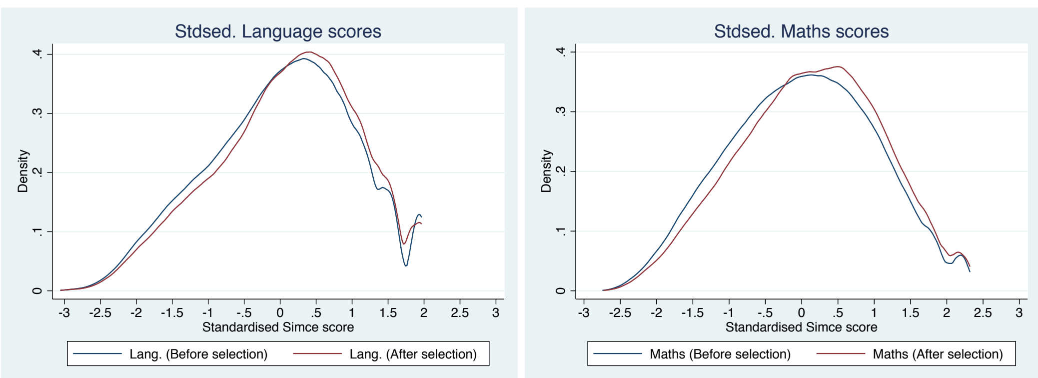


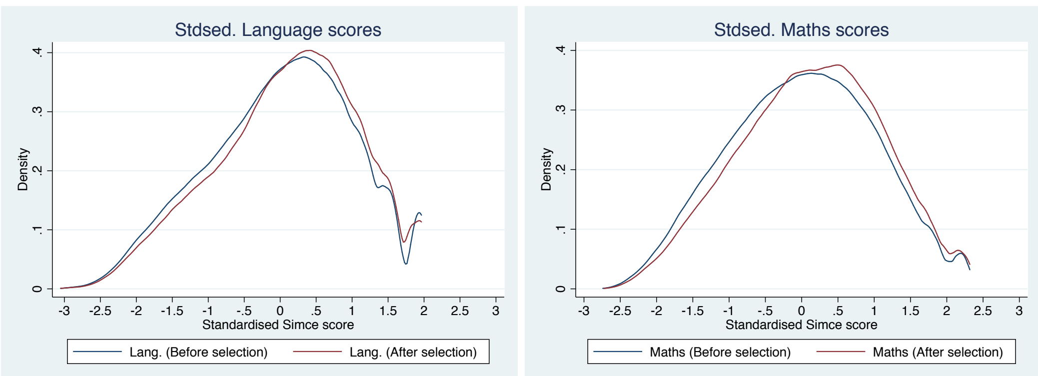
Table 8.20: 4^{th} grade cohort 2009 - Before and after selection

Before selection	Mean	Std. Dev.	Min	Max
Pupil Level				
GPA 4th 2009	5.7	0.9	1	7
GPA 3th 2008	5.8	0.8	1	7
School Language Marks 2009	5.4	0.8	1	7
School Maths Marks 2009	5.3	0.9	1	7
School Language Marks 2008	5.5	0.8	1	7
School Maths Marks 2008	5.4	0.9	1	7
Language Simce Scores 4th 2009	261.9	53.4	99.01	366.7
Maths Simce Scores 4th 2009	252.8	55.1	101.8	380.6
Stdsed. Lang. Simce Score 4th 2009	0.0	1.0	-3.1	2.0
Stdsed. Maths Simce Score 4th 2009	0.0	1.0	-2.7	2.3
Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.3	0.6	8	12
Attendance	91.2	14.5	0	100
Mother Education	3.0	2.0	0	8
Father Education	3.1	2.1	0	8
Household Income	3.0	3.1	0	12
Class Level				
Class size	22.9	14.1	1	49
Peers Average GPA	5.7	0.5	1	7
Teacher Level				
Gender (Female=1)	0.84	0.4	0	1
Years of experience in the system	20.7	13.8	0	40
(Teaching hrs / Contract hrs) Ratio	0.9	0.2	0.03	1
School Level				
Municipal Schools	0.57	0.5	0	1
Private Voucher Schools	0.38	0.5	0	1
Unsubsidised Private Schools	0.05	0.2	0	1
Rural Area	0.46	0.5	0	1
Number of classes per grade	1.34	0.7	1	10
Number of students per grade	30.66	34.1	1	388
Number of students	250,275			
Number of teachers	12,245			
Number of schools	8,164			

After selection	Mean	Std. Dev.	Min	Max
Pupil Level				
GPA 4th 2009	5.9	0.7	1	7
GPA 3th 2008	5.9	0.7	1	7
School Language Marks 2009	5.5	0.7	2.5	7
School Maths Marks 2009	5.4	0.8	2.2	7
School Language Marks 2008	5.6	0.7	1	7
School Maths Marks 2008	5.5	0.8	1	7
Language Simce Scores 4th 2009	267.1	52.7	99.01	366.7
Maths Simce Scores 4th 2009	259.8	54.2	102.6	380.6
Stdsed. Lang. Simce Score 4th 2009	0.1	0.99	-3.1	2.0
Stdsed. Maths Simce Score 4th 2009	0.1	0.98	-2.7	2.3
Gender (Female=1)	0.5	0.5	0.0	1.0
Age	9.2	0.4	8	12
Attendance	92.7	9.6	0	100
Mother Education	3.2	1.9	0	8
Father Education	3.3	2.1	0	8
Household Income	3.1	3.1	0	12
Class Level				
Class size	27.2	6.6	15	45
Peers Average GPA	5.8	0.3	4.148	6.593
Teacher Level				
Gender (Female=1)	0.88	0.3	0	1
Years of experience in the system	17.6	11.9	0	40
(Teaching hrs / Contract hrs) Ratio	0.9	0.2	0.05	1
School Level				
Municipal Schools	0.48	0.5	0	1
Private Voucher Schools	0.45	0.5	0	1
Unsubsidised Private Schools	0.07	0.3	0	1
Rural Area	0.03	0.2	0	1
Number of classes per grade	2.28	0.6	2	5
Number of students per grade	62.16	22.8	30	202
Number of students	54,577			
Number of teachers	2,003			
Number of schools	878			

Notes: (i) The selection sample was made for identification, and the most important sources of dropped observations were schools with less than two classes per grade and schools and schools with specialised teachers only, both sum up to 20% of the original pupil's observations in the cohort. **(ii)** Education level (Mother, Father): (0) Primary Incomplete; (1) Primary Complete; (2) Secondary Incomplete; (3) Secondary Complete; (4) Technical Incomplete; (5) Technical Complete; (6) University Incomplete; (7) University Complete; (8) Postgraduate Studies. **(iii)** Household Income Level (US$ approx. - Dec 2009): (0) Less than $200; (1) From $200 to $400; (2) From $400 to $600; (3) From $600 to $800; (4) From $800 to $1,000; (5) From $1,000 to $1,200; (6) From $1,200 to $1,400; (7) From $1,400 to $2,000; (8) From $2,000 to $2,400; (9) From $2,400 to $2,800; (10) From $2,800 to $3,200; (11) From $3,200 to $3,600; (12) More than $3,600.

Figure 8.24: Kernel distributions: Stdsed. Simce Scores - Before and After Selection
4^{th} grade cohort 2009



Appendix 6.2 Estimation results

Table 8.21: Simce score distribution - 4^{th} Grade cohorts (2005 - 2009)

Simce Score		2005 cohort		2006 cohort		2007 cohort		2008 cohort		2009 cohort	
Percentile		Language	Maths	Language	Maths	Language	Maths	Language	Maths	Language	Maths
10th	*Absolute Score*	183.4	176.3	182.1	175.9	184.4	172.9	187.1	178.4	191.9	185.7
	Standardised Score	-1.36	-1.29	-1.32	-1.29	-1.31	-1.30	-1.37	-1.25	-1.31	-1.22
25th	*Absolute Score*	222.8	213.0	218.1	214.1	219.0	210.3	227.1	212.8	232.0	221.5
	Standardised Score	-0.62	-0.63	-0.66	-0.60	-0.67	-0.63	-0.62	-0.63	-0.56	-0.57
50th	*Absolute Score*	263.8	254.2	259.3	254.7	260.4	251.7	268.5	252.8	271.8	261.8
	Standardised Score	0.15	0.12	0.11	0.12	0.11	0.10	0.15	0.10	0.19	0.16
75th	*Absolute Score*	296.5	291.3	295.6	292.2	296.5	290.6	302.6	291.5	305.5	299.4
	Standardised Score	0.77	0.79	0.78	0.80	0.78	0.79	0.78	0.80	0.82	0.85
90th	*Absolute Score*	323.5	320.1	324.2	322.7	324.8	321.9	330.8	323.2	333.6	329.8
	Standardised Score	1.28	1.31	1.31	1.34	1.31	1.35	1.31	1.38	1.34	1.40
Abs. Mean		258.5	251.1	255.7	251.6	257.4	249.6	264.0	251.9	267.1	259.8
Abs. SD		**52.3**	**54.2**	**53.3**	**54.9**	**52.9**	**55.6**	**53.2**	**54.4**	**52.7**	**54.2**
Stzd. Mean		0.054	0.062	0.044	0.068	0.050	0.064	0.063	0.085	0.097	0.126
Stdz. SD		**0.983**	**0.980**	**0.990**	**0.983**	**0.989**	**0.986**	**0.992**	**0.988**	**0.987**	**0.983**
Number of Schools			1,333		1,312		951		1,123		878
Number of Teachers			3,140		3,045		2,186		2,587		2,003
Number of Students			89,278		89,261		63,579		73,825		54,651

Table 8.22: Estimated Teacher Effects and Hypothetical Treatment Selected sample cohorts (4^{th} grade 2005-2009)

Part A. Estimated of Teacher Effects

		Lambda assigned (λ=0.4) Model 4		*Lambda estimated* Model 1	
		Stdsed. Language score (1)	Stdsed. Maths score (2)	Stdsed. Language score (3)	Stdsed. Maths score (4)
2005	Mean TEs	0.0	0.0	0.0	0.0
	SD TEs	**0.21**	**0.31**	**0.18**	**0.28**
2006	Mean TEs	0.0	0.0	0.0	0.0
	SD TEs	**0.23**	**0.35**	**0.20**	**0.31**
2007	Mean TEs	0.0	0.0	0.0	0.0
	SD TEs	**0.20**	**0.32**	**0.18**	**0.29**
2008	Mean TEs	0.0	0.0	0.0	0.0
	SD TEs	**0.19**	**0.32**	**0.16**	**0.29**
2009	Mean TEs	0.0	0.0	0.0	0.0
	SD TEs	**0.16**	**0.28**	**0.15**	**0.28**

Part B. Hypothetical treatment: Being exposed to 1 SD more effective teacher

		Movement in percentile Ranking (1)	(2)	Movement in percentile Ranking (3)	(4)
2005	*Pupil in the Median*	50th to 59th	50th to 60th	50th to 62nd	50th to 61st
2006	*Pupil in the Median*	50th to 59th	50th to 64th	50th to 58th	50th to 62nd
2007	*Pupil in the Median*	50th to 58th	50th to 64th	50th to 57th	50th to 63rd
2008	*Pupil in the Median*	50th to 58th	50th to 62nd	50th to 57th	50th to 61st
2009	*Pupil in the Median*	50th to 56th	50th to 60th	50th to 56th	50th to 60th

Notes: For Part A: (i) The Mean TEs and SD TEs are obtained from the empirical Bayes distribution. **(ii)** Columns (1) and (3) refers to Standardised Language scores for Model 4 and 1, respectively, while columns (2) and (4) correspond to Standardised Maths scores for Model 4 and 1, respectively. **For Part B: (i)** The pupil in the Median is with respect to the Standardised Simce score for each subject and specific cohort. Those values can be observed from Table: "Simce score distribution - 4^{th} Grade cohorts (2005 - 2009)" in Appendix 2. Here we show the expected movement in the ranking given the hypothetical treatment, assuming no other changes changes in the distribution.

Table 8.23: Estimated of School Effects and Hypothetical Treatment
Selected sample cohorts (4^{th} grade 2005-2009)

Part A. Estimated of School Effects

		Lambda assigned (λ=0.4) **Model 4**		*Lambda estimated* **Model 1**	
		Stdsed. Language score (1)	Stdsed. Maths score (2)	Stdsed. Language score (3)	Stdsed. Maths score (4)
2005	Mean SEs	0.0	0.0	0.0	0.0
	SD SEs	**0.29**	**0.25**	**0.28**	**0.24**
2006	Mean SEs	0.0	0.0	0.0	0.0
	SD SEs	**0.28**	**0.25**	**0.27**	**0.23**
2007	Mean SEs	0.0	0.0	0.0	0.0
	SD SEs	**0.29**	**0.26**	**0.28**	**0.25**
2008	Mean SEs	0.0	0.0	0.0	0.0
	SD SEs	**0.28**	**0.25**	**0.28**	**0.25**
2009	Mean SEs	0.0	0.0	0.0	0.0
	SD SEs	**0.28**	**0.26**	**0.29**	**0.28**

Part B. Hypothetical treatment: Being exposed to 1 SD more effective school

		Movement in percentile Ranking (1)	(2)	Movement in percentile Ranking (3)	(4)
2005	*Pupil in the Median*	50th to 62nd	50th to 60th	50th to 62nd	50th to 59th
2006	*Pupil in the Median*	50th to 61st	50th to 60th	50th to 60th	50th to 59th
2007	*Pupil in the Median*	50th to 61th	50th to 60th	50th to 61th	50th to 60th
2008	*Pupil in the Median*	50th to 61th	50th to 60th	50th to 61th	50th to 60th
2009	*Pupil in the Median*	50th to 61st	50th to 62nd	50th to 60th	50th to 60th

Notes: For Part A: (i) The Mean SEs and SD SEs are obtained from the empirical Bayes distribution. **(ii)** Columns (1) and (3) refers to Standardised Language scores for Model 4 and 1, respectively, while columns (2) and (4) correspond to Standardised Maths scores for Model 4 and 1, respectively. **For Part B: (i)** The pupil in the Median is with respect to the Standardised Simce score for each subject and specific cohort. Those values can be observed from Table: "Simce score distribution - 4^{th} Grade cohorts (2005 - 2009)" in Appendix 2. Here we show the expected movement in the ranking given the hypothetical treatment, assuming no other changes changes in the distribution.

Table 8.24: Empirical Bayes distributions
4^{th} grade 2005 selected sample

Distribution of Teacher Effects (TEs)

	Preset lambda (λ=0.4)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)
1%	-0.504	-0.740	-0.434	-0.662
5%	-0.349	-0.513	-0.307	-0.468
10%	-0.264	-0.388	-0.233	-0.355
25%	-0.139	-0.204	-0.124	-0.189
50%	0.003	0.004	0.004	0.007
75%	0.142	0.209	0.123	0.188
90%	0.264	0.387	0.232	0.354
95%	0.338	0.496	0.296	0.451
99%	0.495	0.726	0.434	0.663
75th - 25th	0.281	0.413	0.247	0.377
90th - 10th	0.528	0.775	0.465	0.709
Mean of TEs	0.000	0.000	0.000	0.000
SD of TEs	**0.209**	**0.307**	**0.183**	**0.279**
Number of Schools				1,333
Number of Teachers				3,140
Number of Students				89,278

Distribution of School Effects (SEs)

	Preset lambda (λ=0.4)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)
1%	-0.703	-0.610	-0.701	-0.609
5%	-0.463	-0.402	-0.451	-0.392
10%	-0.379	-0.329	-0.364	-0.316
25%	-0.191	-0.166	-0.184	-0.160
50%	0.016	0.014	0.015	0.013
75%	0.187	0.162	0.186	0.161
90%	0.355	0.308	0.354	0.308
95%	0.473	0.411	0.466	0.405
99%	0.670	0.582	0.654	0.569
75th - 25th	0.378	0.328	0.369	0.321
90th - 10th	0.734	0.637	0.718	0.624
Mean of SEs	0.001	0.001	0.003	0.003
SD of SEs	**0.285**	**0.248**	**0.280**	**0.244**
Number of Schools				1,333
Number of Teachers				3,140
Number of Students				89,278

Distribution of Student Ability (SA)

	Preset lambda (λ=0.4)		Lambda estimaded	
Percentile	Language (1)	Maths (2)	Language (3)	Maths (4)
1%	-1.111	-1.000	-1.071	-0.945
5%	-0.780	-0.702	-0.742	-0.655
10%	-0.586	-0.527	-0.555	-0.490
25%	-0.278	-0.250	-0.264	-0.233
50%	0.028	0.026	0.024	0.021
75%	0.300	0.270	0.283	0.250
90%	0.539	0.485	0.515	0.455
95%	0.688	0.619	0.658	0.581
99%	0.965	0.868	0.935	0.825
75th - 25th	0.578	0.520	0.547	0.482
90th - 10th	1.124	1.012	1.070	0.944
Mean of SA	0.000	0.000	0.000	0.000
SD of SA	**0.443**	**0.399**	**0.000**	**0.374**
Number of Schools				1,333
Number of Teachers				3,140
Number of Students				89,278

Table 8.25: Empirical Bayes distributions
4^{th} grade 2006 selected sample

Distribution of Teacher Effects (TEs)

	Preset lambda (λ=0.4)		Lambda estimated	
Percentile	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.557	-0.833	-0.482	-0.751
5%	-0.383	-0.573	-0.330	-0.515
10%	-0.292	-0.436	-0.252	-0.392
25%	-0.155	-0.231	-0.131	-0.204
50%	-0.003	-0.005	-0.003	-0.005
75%	0.154	0.231	0.134	0.208
90%	0.301	0.450	0.260	0.404
95%	0.390	0.583	0.327	0.509
99%	0.550	0.823	0.467	0.727
75th - 25th	0.309	0.462	0.264	0.412
90th - 10th	0.593	0.887	0.511	0.797
Mean of TEs	0.000	0.000	0.000	0.000
SD of TEs	**0.234**	**0.349**	**0.199**	**0.310**
Number of Schools				1,312
Number of Teachers				3,045
Number of Students				89,950

Distribution of School Effects (SEs)

	Preset lambda (λ=0.4)		Lambda estimated	
Percentile	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.679	-0.600	-0.630	-0.547
5%	-0.489	-0.432	-0.460	-0.399
10%	-0.377	-0.333	-0.356	-0.309
25%	-0.183	-0.161	-0.179	-0.155
50%	0.005	0.004	0.011	0.009
75%	0.202	0.178	0.195	0.169
90%	0.352	0.311	0.344	0.298
95%	0.447	0.395	0.436	0.378
99%	0.611	0.539	0.584	0.507
75th - 25th	0.384	0.339	0.373	0.324
90th - 10th	0.729	0.644	0.699	0.607
Mean of SEs	0.000	0.000	0.001	0.001
SD of SEs	**0.283**	**0.249**	**0.267**	**0.232**
Number of Schools				1,312
Number of Teachers				3,045
Number of Students				89,950

Distribution of Student Ability (SA)

	Lambda assigned (λ=0.4)		Lambda estimated	
Percentile	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-1.110	-0.917	-1.074	-0.863
5%	-0.768	-0.634	-0.735	-0.591
10%	-0.583	-0.482	-0.556	-0.447
25%	-0.283	-0.234	-0.271	-0.218
50%	0.020	0.016	0.017	0.014
75%	0.301	0.249	0.287	0.230
90%	0.550	0.455	0.530	0.426
95%	0.698	0.576	0.672	0.540
99%	0.977	0.807	0.955	0.767
75th - 25th	0.584	0.483	0.558	0.448
90th - 10th	1.133	0.936	1.086	0.872
Mean of SA	0.000	0.000	0.000	0.000
SD of SA	**0.445**	**0.367**	**0.429**	**0.344**
Number of Schools				1,312
Number of Teachers				3,045
Number of Students				89,950

Table 8.26: Empirical Bayes distributions
4^{th} grade 2007 selected sample

Distribution of Teacher Effects (TEs)

	Preset lambda (λ=0.4)		Lambda estimated	
Percentile	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.477	-0.750	-0.428	-0.696
5%	-0.326	-0.513	-0.298	-0.485
10%	-0.257	-0.404	-0.234	-0.380
25%	-0.136	-0.214	-0.124	-0.202
50%	-0.001	-0.002	0.000	0.000
75%	0.140	0.221	0.125	0.203
90%	0.265	0.416	0.235	0.382
95%	0.334	0.525	0.302	0.492
99%	0.461	0.724	0.408	0.663
75th - 25th	0.277	0.435	0.249	0.404
90th - 10th	0.522	0.820	0.468	0.761
Mean of TEs	0.000	0.000	0.000	0.000
SD of TEs	**0.202**	**0.317**	**0.181**	**0.294**
Number of Schools				951
Number of Teachers				2,186
Number of Students				63,579

Distribution of School Effects (SEs)

	Preset lambda (λ=0.4)		Lambda estimated	
Percentile	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.654	-0.588	-0.644	-0.576
5%	-0.473	-0.425	-0.462	-0.413
10%	-0.376	-0.338	-0.367	-0.328
25%	-0.200	-0.180	-0.196	-0.176
50%	0.003	0.003	0.004	0.004
75%	0.199	0.179	0.202	0.181
90%	0.371	0.334	0.362	0.324
95%	0.455	0.409	0.451	0.403
99%	0.644	0.579	0.634	0.567
75th - 25th	0.399	0.359	0.398	0.356
90th - 10th	0.747	0.672	0.730	0.652
Mean of SEs	-0.001	-0.001	0.000	0.000
SD of SEs	**0.286**	**0.257**	**0.281**	**0.251**
Number of Schools				951
Number of Teachers				2,186
Number of Students				63,579

Distribution of Student Ability (SA)

	Lambda assigned (λ=0.4)		Lambda estimated	
Percentile	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-1.120	-0.968	-1.089	-0.924
5%	-0.789	-0.682	-0.753	-0.639
10%	-0.607	-0.524	-0.576	-0.489
25%	-0.297	-0.257	-0.283	-0.240
50%	0.021	0.018	0.018	0.015
75%	0.312	0.270	0.296	0.251
90%	0.570	0.492	0.544	0.462
95%	0.719	0.622	0.689	0.585
99%	1.018	0.880	0.981	0.833
75th - 25th	0.610	0.527	0.578	0.491
90th - 10th	1.176	1.017	1.120	0.951
Mean of SA	0.000	0.000	0.000	0.000
SD of SA	**0.458**	**0.396**	**0.440**	**0.373**
Number of Schools				951
Number of Teachers				2,186
Number of Students				63,579

Table 8.27: Empirical Bayes distributions
4^{th} grade 2008 selected sample

Distribution of Teacher Effects (TEs)

	Preset lambda (λ=0.4)		Lambda estimated	
Percentile	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.429	-0.723	-0.373	-0.660
5%	-0.312	-0.527	-0.270	-0.478
10%	-0.242	-0.407	-0.213	-0.378
25%	-0.125	-0.210	-0.112	-0.198
50%	0.001	0.001	0.001	0.002
75%	0.126	0.213	0.108	0.192
90%	0.238	0.402	0.207	0.367
95%	0.313	0.527	0.275	0.487
99%	0.436	0.735	0.380	0.674
75th - 25th	0.251	0.423	0.220	0.389
90th - 10th	0.480	0.809	0.420	0.744
Mean of TEs	0.000	0.000	0.000	0.000
SD of TEs	**0.188**	**0.317**	**0.164**	**0.291**
Number of Schools				951
Number of Teachers				2,186
Number of Students				63,579

Distribution of School Effects (SEs)

	Preset lambda (λ=0.4)		Lambda estimated	
Percentile	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-0.709	-0.647	-0.687	-0.625
5%	-0.477	-0.435	-0.463	-0.421
10%	-0.370	-0.338	-0.358	-0.326
25%	-0.173	-0.158	-0.174	-0.158
50%	0.002	0.002	0.003	0.002
75%	0.192	0.175	0.192	0.175
90%	0.353	0.322	0.351	0.320
95%	0.455	0.415	0.464	0.422
99%	0.599	0.546	0.604	0.550
75th - 25th	0.365	0.333	0.367	0.334
90th - 10th	0.723	0.659	0.709	0.645
Mean of SEs	0.001	0.001	0.003	0.003
SD of SEs	**0.279**	**0.255**	**0.276**	**0.251**
Number of Schools				951
Number of Teachers				2,186
Number of Students				63,579

Distribution of Student Ability (SA)

	Preset lambda (λ=0.4)		Lambda estimated	
Percentile	Language (3)	Maths (4)	Language (5)	Maths (6)
1%	-1.101	-0.880	-1.068	-0.825
5%	-0.780	-0.623	-0.749	-0.578
10%	-0.598	-0.477	-0.571	-0.441
25%	-0.298	-0.238	-0.282	-0.218
50%	0.018	0.015	0.016	0.012
75%	0.310	0.248	0.294	0.227
90%	0.567	0.453	0.540	0.417
95%	0.724	0.578	0.695	0.536
99%	1.023	0.817	0.992	0.766
75th - 25th	0.609	0.486	0.576	0.445
90th - 10th	1.165	0.930	1.111	0.858
Mean of SA	0.000	0.000	0.000	0.000
SD of SA	**0.456**	**0.364**	**0.438**	**0.339**
Number of Schools				951
Number of Teachers				2,186
Number of Students				63,579

Table 8.28: Empirical Bayes distributions
4^{th} grade 2009 selected sample

Distribution of Teacher Effects (TEs)

Percentile	Preset lambda (λ=0.4) Language (3)	Preset lambda (λ=0.4) Maths (4)	Lambda estimated Language (5)	Lambda estimated Maths (6)
1%	-0.346	-0.596	-0.350	-0.654
5%	-0.256	-0.441	-0.244	-0.457
10%	-0.193	-0.332	-0.192	-0.360
25%	-0.106	-0.183	-0.100	-0.188
50%	0.000	0.000	-0.001	-0.001
75%	0.105	0.181	0.105	0.196
90%	0.215	0.370	0.193	0.361
95%	0.285	0.492	0.246	0.461
99%	0.394	0.680	0.349	0.654
75th - 25th	0.211	0.364	0.205	0.384
90th - 10th	0.408	0.703	0.385	0.720
Mean of TEs	0.005	0.008	0.000	0.000
SD of TEs	**0.162**	**0.279**	**0.150**	**0.280**
Number of Schools				951
Number of Teachers				2,186
Number of Students				63,579

Distribution of School Effects (SEs)

Percentile	Preset lambda (λ=0.4) Language (3)	Preset lambda (λ=0.4) Maths (4)	Lambda estimated Language (5)	Lambda estimated Maths (6)
1%	-0.693	-0.649	-0.756	-0.726
5%	-0.481	-0.450	-0.479	-0.461
10%	-0.368	-0.345	-0.386	-0.371
25%	-0.187	-0.175	-0.176	-0.169
50%	0.021	0.020	0.016	0.015
75%	0.188	0.176	0.198	0.190
90%	0.343	0.322	0.347	0.334
95%	0.450	0.422	0.460	0.442
99%	0.600	0.562	0.681	0.654
75th - 25th	0.375	0.351	0.373	0.359
90th - 10th	0.712	0.667	0.734	0.705
Mean of SEs	0.000	0.000	0.003	0.003
SD of SEs	**0.279**	**0.262**	**0.289**	**0.277**
Number of Schools				951
Number of Teachers				2,186
Number of Students				63,579

Distribution of Student Ability (SA)

Percentile	Preset lambda (λ=0.4) Language (3)	Preset lambda (λ=0.4) Maths (4)	Lambda estimated Language (5)	Lambda estimated Maths (6)
1%	-1.118	-0.893	-1.150	-0.907
5%	-0.771	-0.616	-0.791	-0.624
10%	-0.590	-0.471	-0.607	-0.479
25%	-0.287	-0.229	-0.297	-0.234
50%	0.020	0.016	0.022	0.017
75%	0.307	0.245	0.318	0.251
90%	0.554	0.443	0.571	0.451
95%	0.700	0.559	0.722	0.569
99%	0.973	0.777	1.002	0.790
75th - 25th	0.594	0.474	0.614	0.485
90th - 10th	1.144	0.914	1.179	0.930
Mean of SA	0.000	0.000	0.000	0.000
SD of SA	**0.448**	**0.358**	**0.461**	**0.364**
Number of Schools				951
Number of Teachers				2,186
Number of Students				63,579

Appendix 6.3 Distribution of teachers by type of teacher specialisation

Figure 8.25: Disaggregated type of teacher specialisation 2005 - 2009

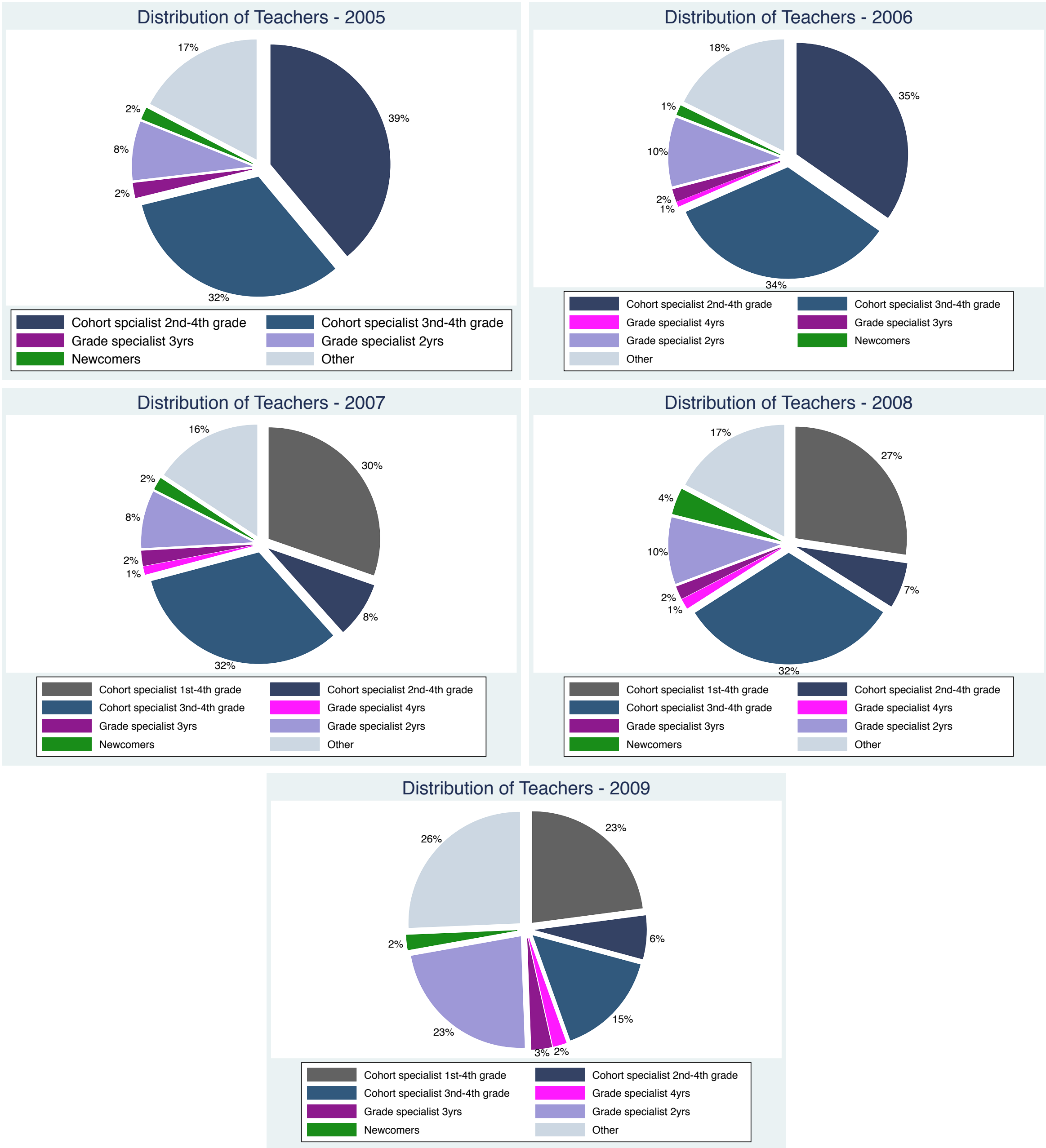


Figure 8.26: Aggregated type of teacher specialisation 2005 - 2009
Municipal schools

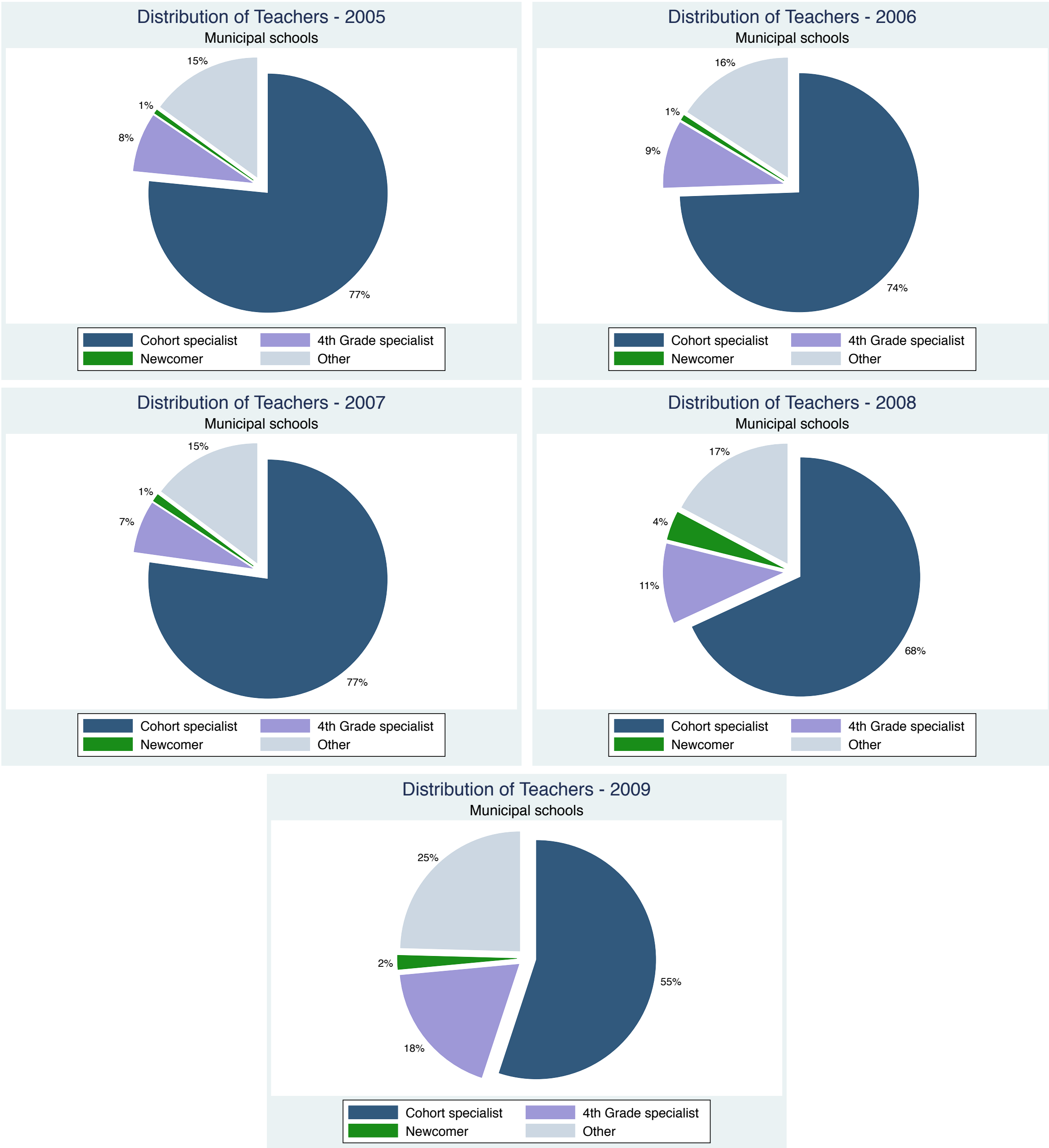


Figure 8.27: Aggregated type of teacher specialisation 2005 - 2009
Private Voucher schools

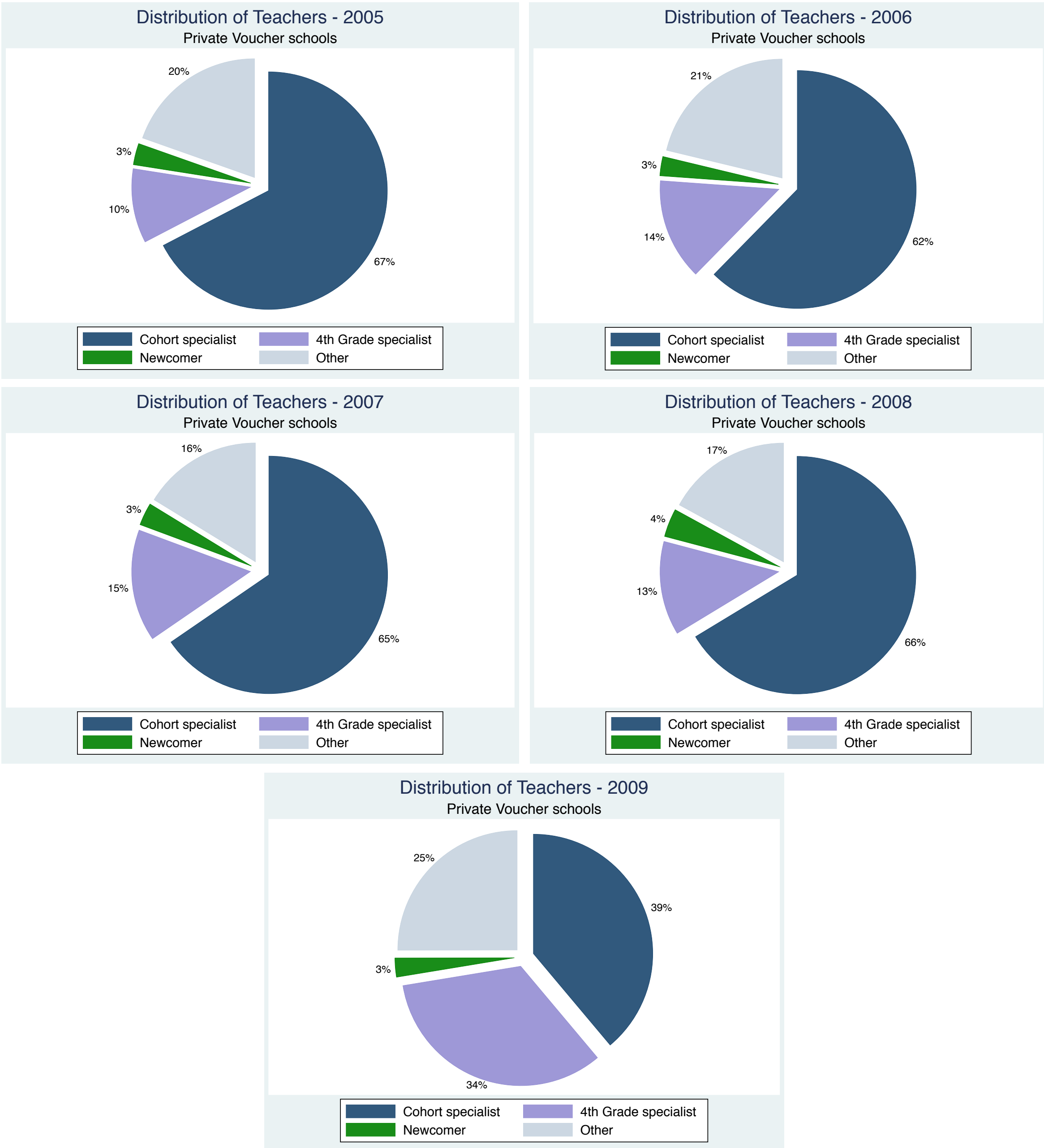


Figure 8.28: Aggregated teacher specialisation 2005 - 2009
Unsubsidised Private schools

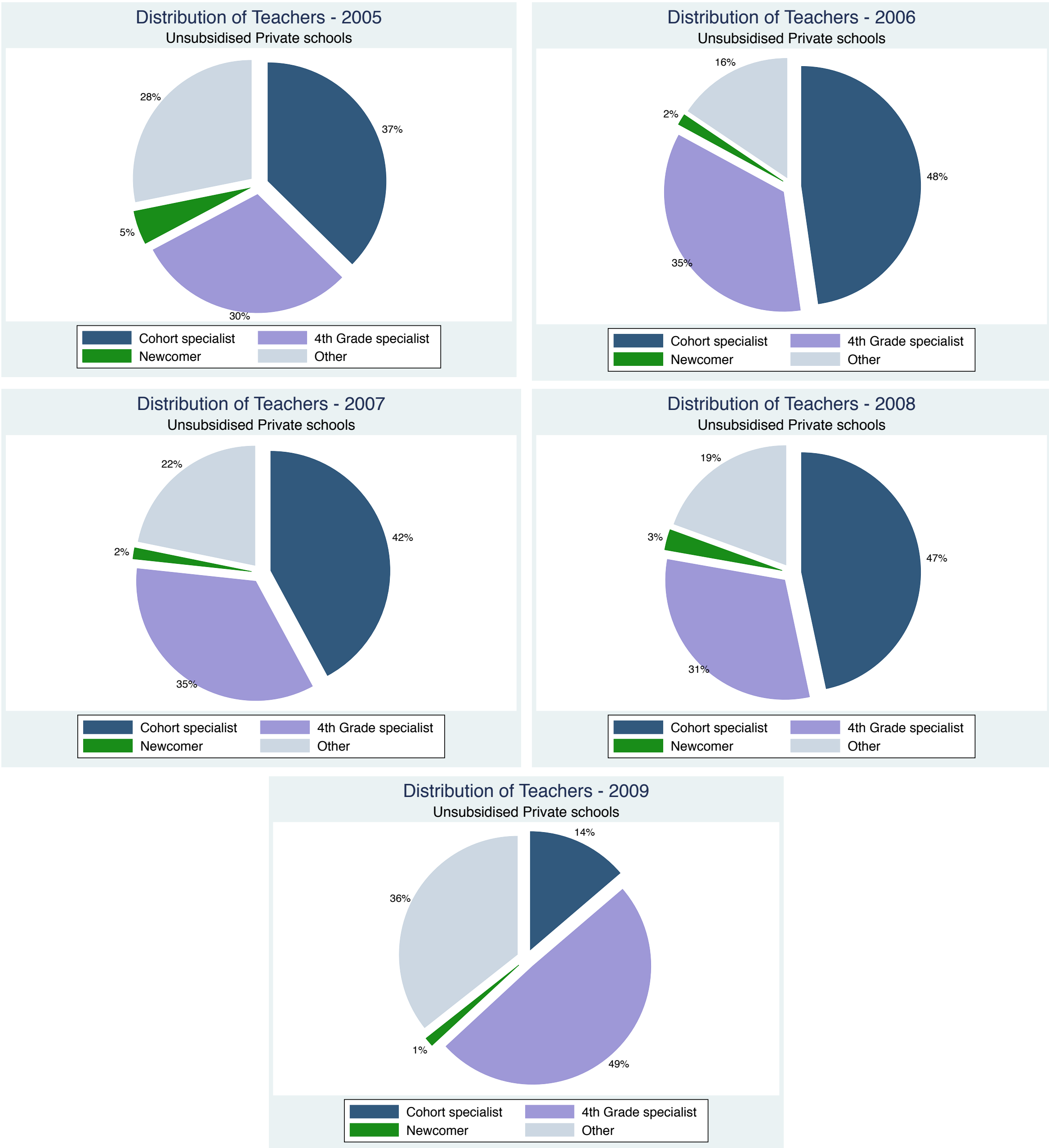


Figure 8.29: Disaggregated teacher specialisation histogram 2005 - 2009 by teacher quality

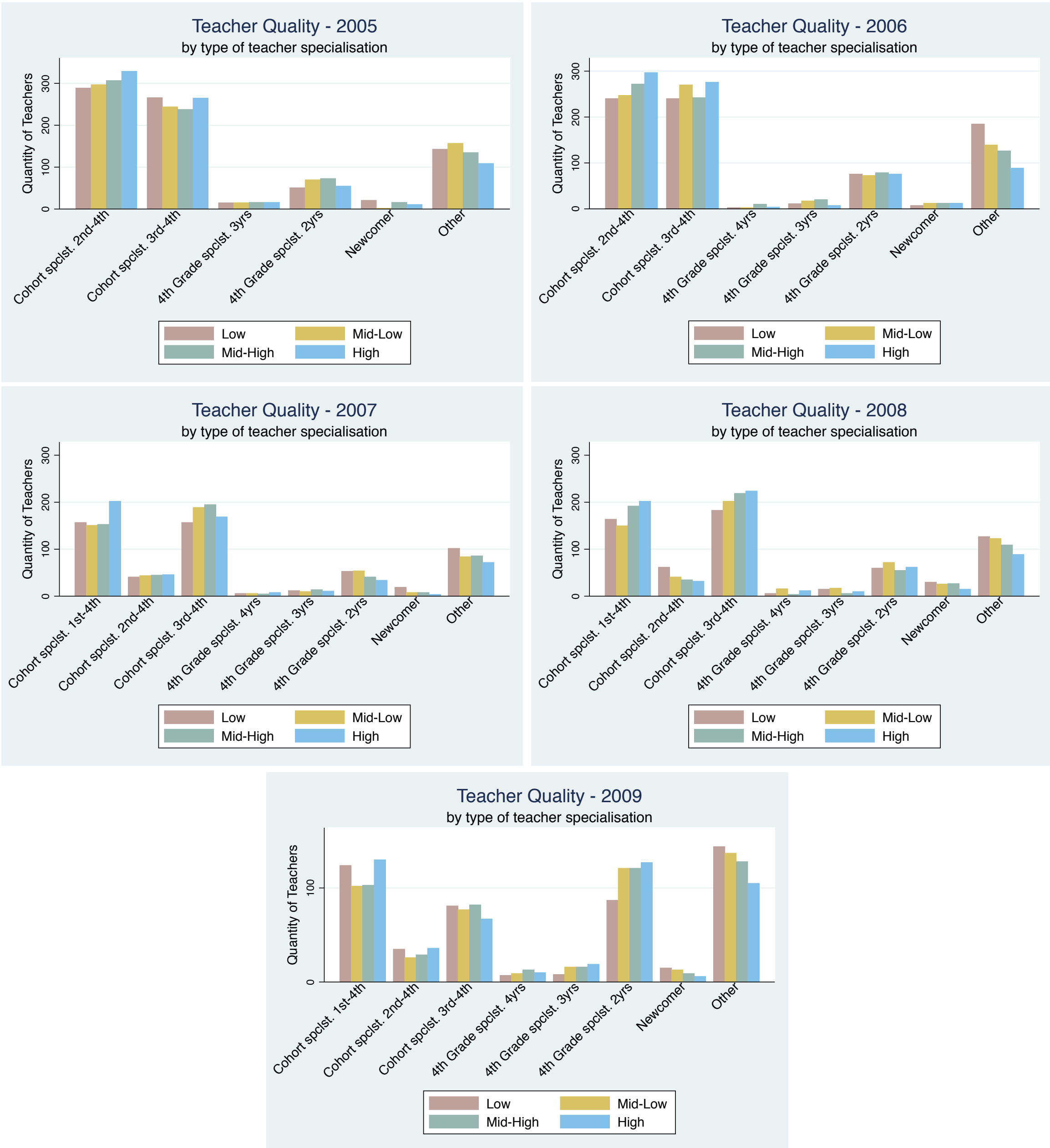


Figure 8.30: Disaggregated teacher specialisation distribution 2005 - 2009 by teacher quality

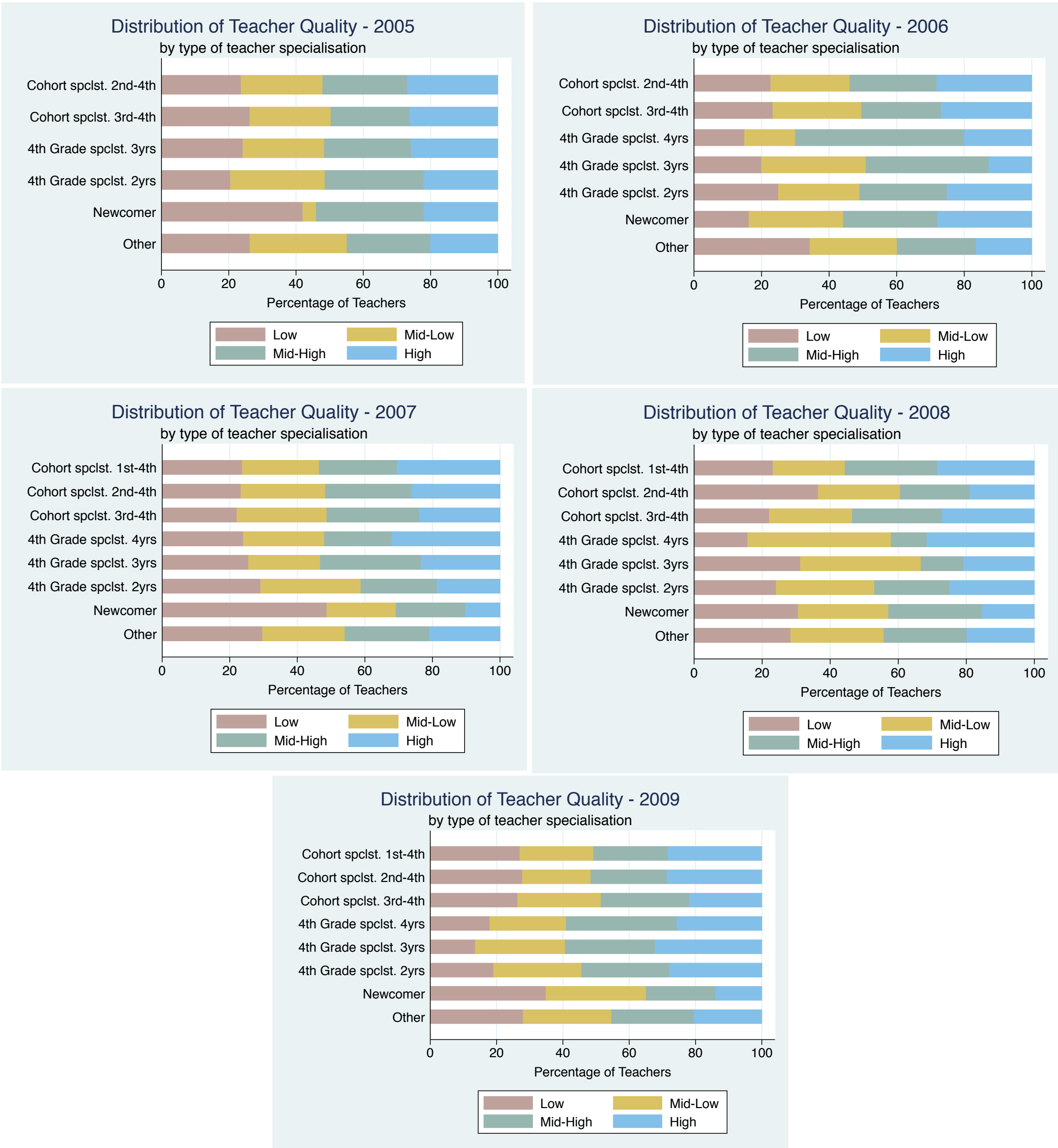
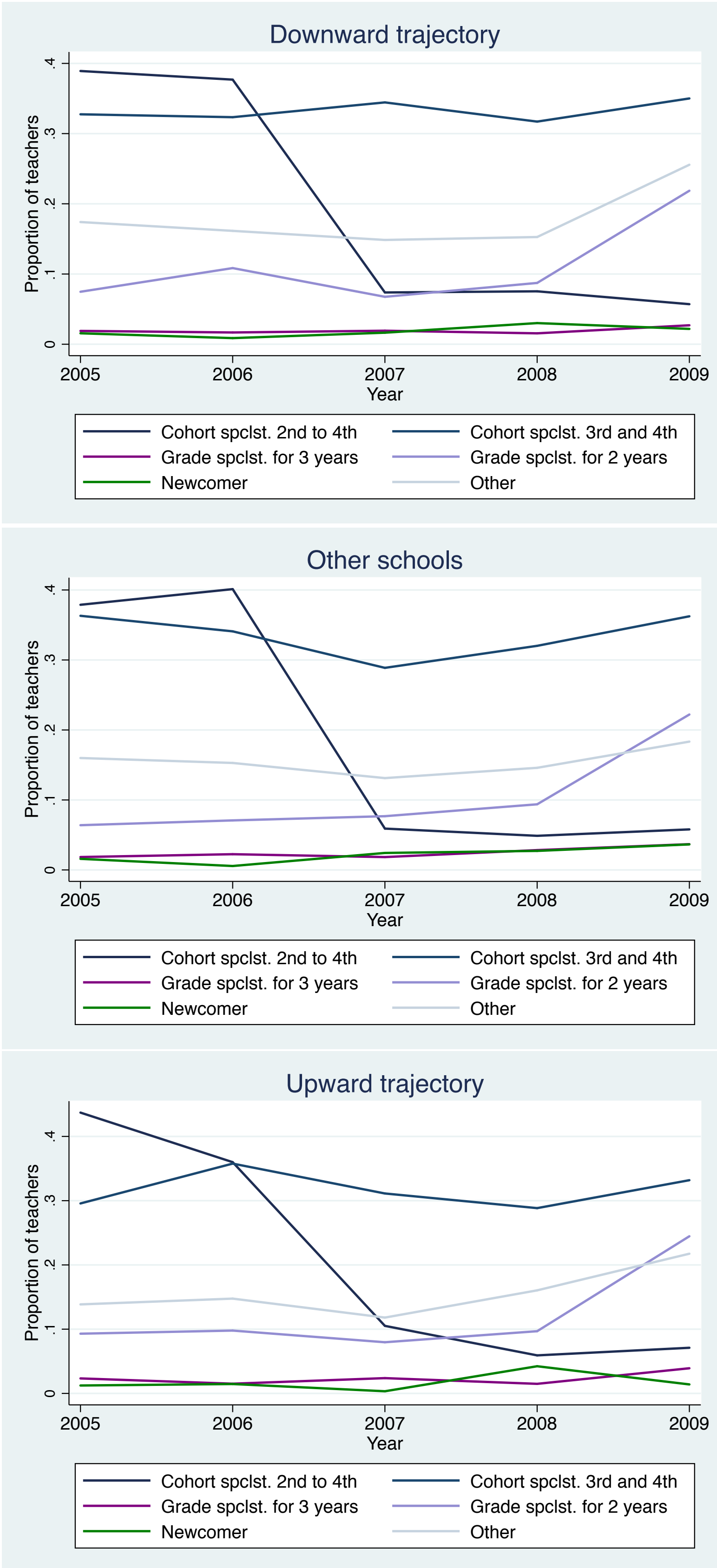


Figure 8.31: Disaggregated teacher specialisation evolution 2005 - 2009 by school quality trajectory



Appendix 6.4 Descriptive statistics - School level

Table 8.29: Cross Section cohorts 4^{th} Grade 2005 - 2009

	2005				2006				2007				2008				2009			
Observed variables	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
School characteristics																				
Average Stdz Languge Simce Score (4th)	0.02	0.5	-1.22	1.34	0.01	0.4	-1.19	1.22	0.02	0.5	-1.24	1.48	0.0	0.5	-1.26	1.44	0.06	0.5	-1.16	1.26
Average Stdz Maths Simce Score (4th)	0.03	0.5	-1.35	1.44	0.03	0.5	-1.43	1.27	0.03	0.5	-1.17	1.47	0.0	0.5	-1.22	1.57	0.08	0.5	-1.24	1.41
Municipal Schools	0.59	0.5	0	1	0.58	0.5	0	1	0.58	0.5	0	1	0.6	0.5	0	1	0.48	0.5	0	1
Private Voucher Schools	0.36	0.5	0	1	0.36	0.5	0	1	0.36	0.5	0	1	0.4	0.5	0	1	0.45	0.5	0	1
Unsubsidised Private Schools	0.05	0.2	0	1	0.06	0.2	0	1	0.06	0.2	0	1	0.1	0.2	0	1	0.07	0.3	0	1
Rural Area	0.04	0.2	0	1	0.04	0.2	0	1	0.03	0.2	0	1	0.0	0.2	0	1	0.03	0.2	0	1
Number of teachers (classes) per grade	2.36	0.6	2	5	2.32	0.6	2	5	2.30	0.6	2	5	2.3	0.6	2	5	2.28	0.6	2	5
Number of students per grade	67.0	25.2	30	209	68.0	26.1	31	206	66.9	25.2	32	193	65.7	25.4	30	204	62.2	22.8	30	202
School Quality (SE quartiles)	2.50	1.1	1	4	2.50	1.1	1	4	2.50	1.1	1	4	2.5	1.1	1	4	2.50	1.1	1	4
EB Language School Effects (SEs)	0.00	0.29	-0.96	0.99	0.00	0.28	-0.85	0.75	0.00	0.29	-0.99	1.39	0.0	0.28	-0.91	1.07	0.00	0.29	-0.92	0.94
EB Maths School Effects (SEs)	0.00	0.25	-0.83	0.86	0.00	0.25	-0.75	0.66	0.00	0.26	-0.89	1.25	0.0	0.26	-0.83	0.98	0.00	0.28	-0.88	0.90
Average EB Language Teacher Effects (TEs)	0.00	0.16	-0.52	0.53	0.00	0.17	-0.72	0.60	0.00	0.15	-0.48	0.61	0.0	0.14	-0.53	0.49	0.00	0.11	-0.37	0.38
Average EB Maths Teacher Effects (TEs)	0.00	0.23	-0.76	0.78	0.00	0.26	-1.08	0.90	0.00	0.24	-0.76	0.97	0.0	0.24	-0.90	0.83	0.00	0.21	-0.69	0.71
Teacher characteristics (school average)																				
Gender (Female=1)	0.9	0.2	0	1	0.9	0.2	0	1	0.9	0.2	0	1	0.9	0.2	0	1	0.9	0.2	0	1
Age	48.5	8.4	25.5	79.5	48.2	8.4	23.3	76.5	48.6	8.6	24.0	69.0	46.8	8.5	24.0	87.5	45.8	8.5	26.0	83.0
Years of experience in the system	20.0	9.5	0.0	40.0	20.4	9.6	0.0	40.0	20.6	9.9	1.0	40.0	18.9	9.5	0.0	40.0	17.6	9.4	0.0	38.5
(Teaching hrs / Contract hrs) Ratio	0.9	0.1	0.2	1	0.9	0.1	0.2	1	0.9	0.1	0.2	1	0.9	0.1	0.1	1	0.9	0.1	0.0	1
Additional qualitfications	0.4	0.4	0.0	1	0.4	0.4	0.0	1					0.5	0.4	0.0	1	0.5	0.4	0.0	1
Post-graduate studies	0.0	0.1	0.0	1	0.0	0.1	0.0	1					0.0	0.1	0.0	1	0.0	0.1	0.0	1
Expect students complete technical degree	0.2	0.3	0.0	1	0.2	0.3	0.0	1	0.3	0.3	0.0	1	0.2	0.3	0.0	1	0.0	0.0	0.0	0
Expect students complete university degree	0.2	0.4	0.0	1	0.2	0.4	0.0	1	0.3	0.4	0.0	1	0.3	0.4	0.0	1	0.0	0.0	0.0	0
Cohorts specialist	0.72	0.35	0.0	1	0.68	0.37	0.0	1	0.71	0.35	0.0	1	0.66	0.35	0.0	1	0.65	0.37	0.0	1
Grade specialist	0.10	0.22	0.0	1	0.12	0.24	0.0	1	0.10	0.22	0.0	1	0.11	0.22	0.0	1	0.25	0.33	0.0	1
Newcomers	0.02	0.09	0.0	1	0.01	0.08	0.0	1	0.02	0.09	0.0	1	0.04	0.14	0.0	1	0.02	0.10	0.0	1
Other training	0.17	0.27	0.0	1	0.18	0.29	0.0	1	0.15	0.26	0.0	1	0.17	0.28	0.0	1	0.25	0.32	0.0	1
Low quality Teacher	0.25	0.31	0.0	1	0.25	0.31	0.0	1	0.25	0.31	0.0	1	0.25	0.32	0.0	1	0.25	0.32	0.0	1
Mid-Low quality Teacher	0.25	0.30	0.0	1	0.25	0.30	0.0	1	0.25	0.31	0.0	1	0.25	0.30	0.0	1	0.25	0.31	0.0	1
Mid-High quality Teacher	0.25	0.29	0.0	1	0.25	0.30	0.0	1	0.25	0.30	0.0	1	0.25	0.29	0.0	1	0.25	0.31	0.0	1
High quality Teacher	0.25	0.31	0.0	1	0.24	0.31	0.0	1	0.25	0.33	0.0	1	0.25	0.32	0.0	1	0.25	0.32	0.0	1
Principal characteristics																				
Gender (Female=1)	0.5	0.5	0	1	0.5	0.5	0	1	0.5	0.5	0	1	0.5	0.5	0	1	0.5	0.5	0	1
Age	58.6	10.6	25	85	58.7	10.7	26	85	58.6	10.7	28	85	58.0	10.5	29	85	58.1	10.8	30	85
Years of experience in the system	30.0	10.8	0	63	30.3	10.4	0	65	30.7	10.5	0	66	29.8	10.5	0	67	29.8	10.2	0	67
(Teaching hrs / Contract hrs) Ratio	0.5	0.5	0.00	1	0.4	0.5	0.00	1	0.4	0.5	0.00	1	0.4	0.5	0.00	1	0.3	0.5	0.00	1
Others																				
Without teacher variables (staff data base)	0.0	0.0	0	0	0.0	0.0	0	0	0.0	0.0	0	0	0.0	0.0	0	0	0.0	0.0	0	0
Without teacher variables (questionnaire)	0.0	0.1	0	1	0.0	0.1	0	1	0.0	0.1	0	1	0.0	0.1	0	1	0.1	0.2	0	1
Without principal variables (staff data base)	0.1	0.2	0	1	0.1	0.3	0	1	0.1	0.2	0	1	0.1	0.2	0	1	0.1	0.2	0	1
Number of Students	89,278				89,261				63,579				73,825				54,651			
Number of Teachers	3,140				1,312				951				1,123				878			
Number of Schools	1,333				3,045				2,186				2,587				2,003			

Notes: **(i)** The descriptive statistics of each variable correspond the average across schools given the average observed within school in 4th grade cohorts over the period 2005 - 2009. **(ii)** The dummy variables show the proportion of each category in both school panels (Municipal, Private Voucher, Unsubsidised Private schools; Rural Area; Teacher's gender; Additional teacher qualifications; Post-graduate studies of teachers; Expectations on student completion; Cohort specialist; Grade specialist; Newcomers; Other specialisation; Low, Mid-Low, Mid-High, High quality teachers; Principal's gender).

Appendix 6.5 Selection sample of trackable schools

Table 8.30: Transition matrices - Groups 1.1 & 1.2

Transitions of School Quality
by School Effects (SE)
Group 1.1 (11111)

		2006				2007				2008					2009				
		Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
2005	Low	24	17	7	2	27	15	7	1	28	12	8	2	0	24	19	6	1	0
2005	Mid-Low	20	24	8	14	16	23	21	6	17	21	21	7	0	17	15	19	15	0
2005	Mid-High	18	14	27	26	13	22	25	25	21	11	27	26	0	13	23	29	20	0
2005	High	3	10	28	54	1	14	32	48	2	19	25	49	0	6	13	27	49	0
		65	65	70	96	57	74	85	80	68	63	81	84	0	60	70	81	85	0
		Total 2005-06		**296**		**Total 2007**		**296**		**Total 2008**		**296**			**Total 2009**		**296**		

Transitions of School Quality (percentage)
by School Effects (SE)
Group 1.1 (11111)

		2006				2007				2008					2009				
		Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
2005	Low	37%	26%	10%	2%	47%	20%	8%	1%	41%	19%	10%	2%		40%	27%	7%	1%	
2005	Mid-Low	31%	37%	11%	15%	28%	31%	25%	8%	25%	33%	26%	8%		28%	21%	23%	18%	
2005	Mid-High	28%	22%	39%	27%	23%	30%	29%	31%	31%	17%	33%	31%		22%	33%	36%	24%	
2005	High	5%	15%	40%	56%	2%	19%	38%	60%	3%	30%	31%	58%		10%	19%	33%	58%	

Transitions of School Quality
by School Effects (SE)
Group 1.2 (11X11)

		2006				2007				2008					2009				
		Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
2005	Low	14	9	7	2	0	0	0	0	10	10	10	2	0	15	8	6	3	0
2005	Mid-Low	8	10	5	5	0	0	0	0	12	9	7	0	0	13	6	7	2	0
2005	Mid-High	4	7	9	9	0	0	0	0	5	9	8	7	0	5	5	12	7	0
2005	High	3	6	8	20	0	0	0	0	3	8	9	17	0	2	12	8	15	0
		29	32	29	36	0	0	0	0	30	36	34	26		35	31	33	27	
		Total 2005-06		**126**		**Total 2007**		**0**		**Total 2008**		**126**			**Total 2009**		**126**		

Transitions of School Quality (percentage)
by School Effects (SE)
Group 1.2 (11X11)

		2006				2007				2008					2009				
		Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
2005	Low	48%	28%	24%	6%					33%	28%	29%	8%		43%	26%	18%	11%	
2005	Mid-Low	28%	31%	17%	14%					40%	25%	21%	0%		37%	19%	21%	7%	
2005	Mid-High	14%	22%	31%	25%					17%	25%	24%	27%		14%	16%	36%	26%	
2005	High	10%	19%	28%	56%					10%	22%	26%	65%		6%	39%	24%	56%	

Notes: **(i)** Groups 1.1 and 1.2 correspond to schools which were observed until the end of the panel (2009). Schools in Group 1.2 have one missing observation in 2007, but we include them in the reduced school panel (RSP) because we are aware of the lack of individual data for this particular year. **(ii)** The total number of schools in Group 1.1 is 296, while in Group 1.2 is 126. **(iii)** In 2008 and 2009 we identify those schools who left the RSP because they shifted to a subject specialist (SS) teacher scheme, and we can not estimate TEs and SEs for them. **(iv)** All transitions are with respect to the base year 2005; the lower triangle (red-light colour) represents downward transitions, which we also include Low-Low combination as it cannot move to a lower category; the upper triangle (blue-light colour) show upward movements, and we also consider the case High-High as there is not a higher quality level to move.

Table 8.31: Transition matrices - Groups 2.1, 2.2 & 2.3

Transitions of School Quality
by School Effects (SE)
Group 2.1 (1111X)

	2006				2007				2008					2009				
2005	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
Low	20	14	5	4	20	8	12	3	20	13	3	7	0	0	0	0	0	13
Mid-Low	11	18	8	4	15	12	11	3	13	9	12	7	0	0	0	0	0	12
Mid-High	7	7	15	14	7	13	12	11	8	10	17	8	0	0	0	0	0	14
High	4	4	14	18	3	8	15	14	1	9	9	21	0	0	0	0	0	20
	42	43	42	40	45	41	50	31	42	41	41	43	0	0	0	0	0	59
	Total 2005-06		167		Total 2007		167		Total 2008		167			Total 2009		0		

Transitions of School Quality (percentage)
by School Effects (SE)
Group 2.1 (1111X)

	2006				2007				2008					2009				
2005	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
Low	48%	33%	12%	10%	44%	20%	24%	10%	48%	32%	7%	16%						22%
Mid-Low	26%	42%	19%	10%	33%	29%	22%	10%	31%	22%	29%	16%						20%
Mid-High	17%	16%	36%	35%	16%	32%	24%	35%	19%	24%	41%	19%						24%
High	10%	9%	33%	45%	7%	20%	30%	45%	2%	22%	22%	49%						34%

Transitions of School Quality
by School Effects (SE)
Group 2.2 (11X1X)

	2006				2007				2008					2009				
2005	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
Low	9	6	2	1	0	0	0	0	10	6	0	2	0	0	0	0	0	4
Mid-Low	2	4	6	7	0	0	0	0	1	6	8	4	0	0	0	0	0	7
Mid-High	3	11	7	7	0	0	0	0	6	7	10	5	0	0	0	0	0	11
High	1	5	6	12	0	0	0	0	2	8	6	8	0	0	0	0	0	5
	15	26	21	27	0	0	0	0	19	27	24	19	0	0	0	0	0	27
	Total 2005-06		89		Total 2007		0		Total 2008		89			Total 2009		0		

Transitions of School Quality (percentage)
by School Effects (SE)
Group 2.2 (11X1X)

	2006				2007				2008					2009				
2005	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
Low	21%	14%	5%	3%					24%	15%	0%	5%						7%
Mid-Low	5%	9%	14%	18%					2%	15%	20%	9%						12%
Mid-High	7%	26%	17%	18%					14%	17%	24%	12%						19%
High	2%	12%	14%	30%					5%	20%	15%	19%						8%

Transitions of School Quality
by School Effects (SE)
Group 2.3 (111XX)

	2006				2007				2008					2009				
2005	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
Low	12	7	5	2	12	10	3	1	0	0	0	0	8	0	0	0	0	2
Mid-Low	8	5	5	2	9	6	1	4	0	0	0	0	11	0	0	0	0	1
Mid-High	6	5	13	7	5	10	10	6	0	0	0	0	12	0	0	0	0	2
High	1	4	3	12	2	1	8	9	0	0	0	0	10	0	0	0	0	1
	27	21	26	23	28	27	22	20	0	0	0	0	41	0	0	0	0	6
	Total 2005-06		97		Total 2007		97		Total 2008		0			Total 2009		0		

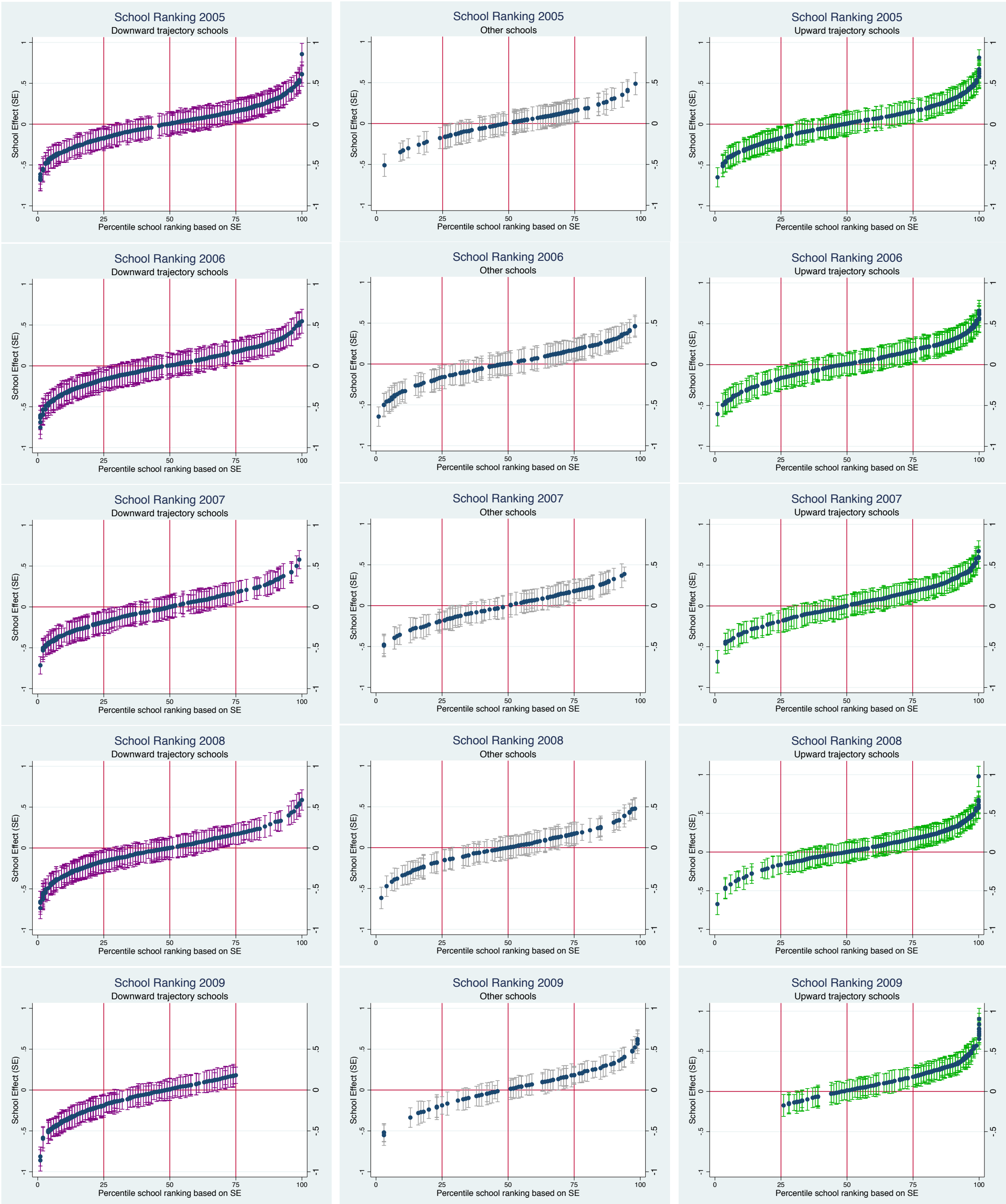
Transitions of School Quality (percentage)
by School Effects (SE)
Group 2.3 (111XX)

	2006				2007				2008					2009				
2005	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	Low	Mid-Low	Mid-High	High	SS scheme	Low	Mid-Low	Mid-High	High	SS scheme
Low	44%	33%	19%	9%	43%	37%	14%	5%					20%					33%
Mid-Low	30%	24%	19%	9%	32%	22%	5%	20%					27%					17%
Mid-High	22%	24%	50%	30%	18%	37%	45%	30%					29%					33%
High	4%	19%	12%	52%	7%	4%	36%	45%					24%					17%

Notes: (i) Groups 2.1, 2.2 and 2.3 correspond to schools which disappear from the reduced school panel (RSP) either in 2008 or 2009. Schools in Group 2.2 have one missing observation in 2007 and disappear in 2009, but we include them in the RSP because we are aware of the lack of individual data for this particular year. **(ii)** The total number of schools in Group 2.1 is 167; Group 2.2 89; and Group 2.3 97. **(iii)** In 2008 and 2009 we identify those schools who left the RSP because they shifted to a subject specialist (SS) teacher scheme, and we can not estimate TEs and SEs for them. **(iv)** All transitions are with respect to the base year 2005; the lower triangle (red-light colour) represents downward transitions, which we also include Low-Low combination as it cannot move to a lower category; the upper triangle (blue-light colour) show upward movements, and we also consider the case High-High as there is not a higher quality level to move.

Appendix 6.6 School rankings 2005 - 2009

Figure 8.32: Caterpillar plots by quality trajectory group



Notes: **(i)** School rankings based on estimated School Effects (SEs), where the lowest percentile (1^{st}) is for the 1st percent lower SE schools, and the 100^{th} percentile corresponds to the 1 percent highest estimated SE schools. **(ii)** The percentile ranking is separated in quartiles, which we call the school quality categories: Low, Mid-Low, Mid-High and High. **(iii)** Whiskers represent the standard errors of SE estimates.